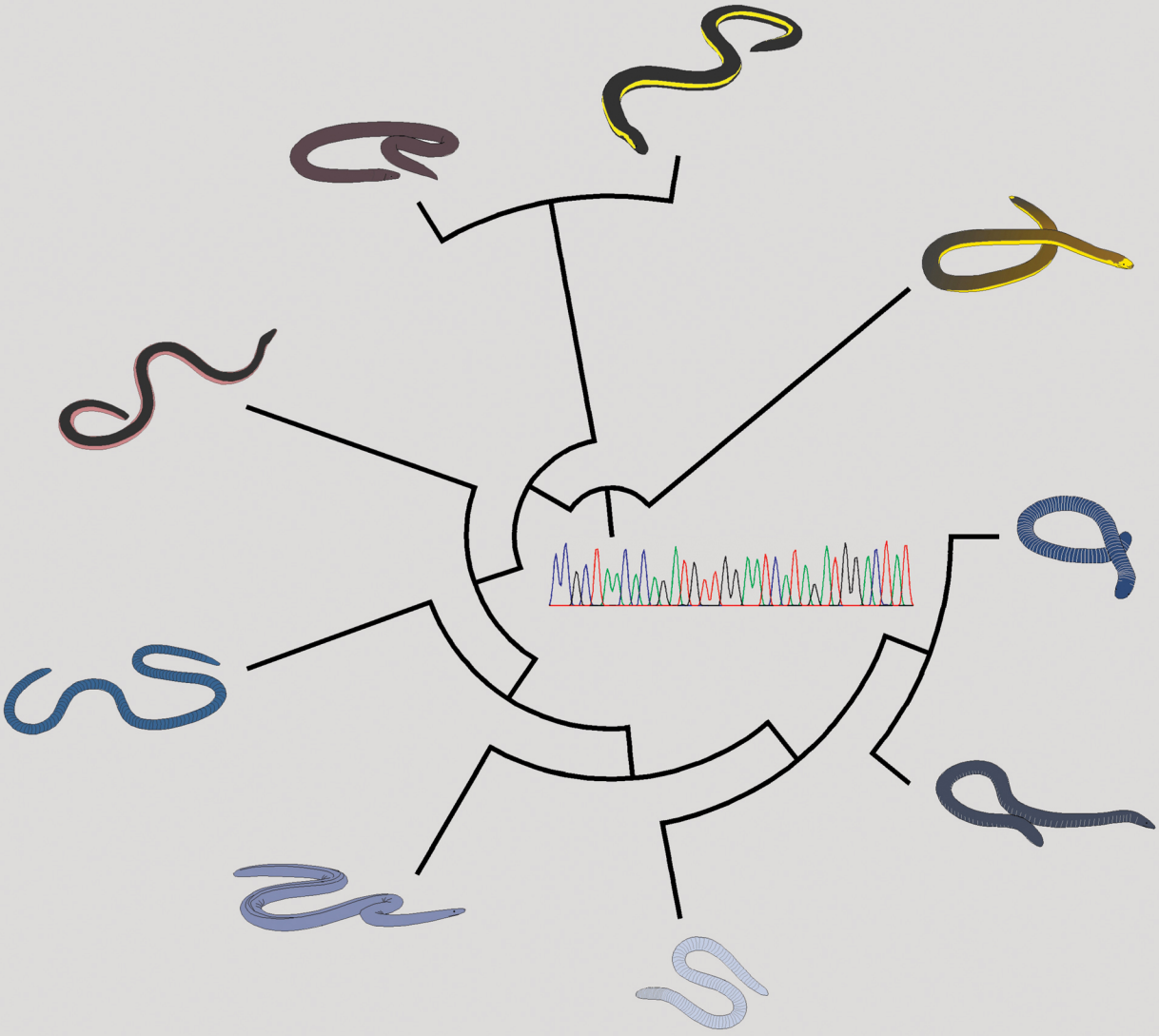
UNIVERSIDAD AUTONOMA DE MADRID
DEPARTAMENTO DE BIOLOGÍA
FACULTAD DE CIENCIAS

CSIC
DEPARTAMENTO DE BIODIVERSIDAD Y BIOLOGÍA EVOLUTIVA
MUSEO NACIONAL DE CIENCIAS NATURALES

PH.D. THESIS

Molecular evolution and phylogenetics of caecilian amphibians (Gymnophiona)



Diego San Mauro Martín

OCTOBER 2006

UNIVERSIDAD AUTONOMA DE MADRID

DEPARTAMENTO DE BIOLOGÍA
FACULTAD DE CIENCIAS

MINISTERIO DE EDUCACIÓN Y CIENCIA

CSIC
Consejo Superior de Investigaciones Científicas

DEPARTAMENTO DE BIODIVERSIDAD Y BIOLOGÍA EVOLUTIVA
MUSEO NACIONAL DE CIENCIAS NATURALES

Molecular evolution and phylogenetics of caecilian amphibians (Gymnophiona)

Memoria presentada por DIEGO SAN MAURO MARTÍN para optar al Grado de Doctor Europeo en Ciencias Biológicas

Diego San Mauro Martín

Vº Bº Director de Tesis

Dr. Rafael Zardoya San Sebastián

Vº Bº Tutor de Tesis

Prof. Arturo Morales Muñíz

Madrid, Octubre 2006

To my parents and brothers.

A mis padres y hermanos.

ACKNOWLEDGEMENTS

First, I thank my advisor, Rafa Zardoya, for his continuous support and enthusiastic supervision of my scientific career. I could not have imagined having a better advisor and mentor for my Ph.D. He showed me different ways to approach a research problem and the need to be persistent to accomplish any goal. I have learnt from him much more than I could have ever imagined. So, thank you very much Rafa!

I would like to gratefully acknowledge Lukas Rüber for helping me so much during the first steps of my Ph.D. He really acted as a "minor" advisor, guiding my study, and being always there to listen and to give advice. I do think he is a model researcher, and model person too. Thanks a lot my friend!

A special thanks goes to Dave Gower and Mark Wilkinson, who I consider co-advisors of my scientific career. They are most responsible for really getting me in touch with the world of caecilian amphibians, particularly during the many visits I have done to their lab in London. Perhaps more importantly, they taught me how to think differently to explore novel ideas, and to ask myself good questions (sometimes even philosophical ones) to help me wisely think through everything.

Besides Dave and Mark, I also particularly thank all other members of the "Caecilian Gang": Hendrik, Alex, James, Simon, Biju, and Sam. They kindly "adopted" me while in London, and really let me feel as part of the group there (…almost like a family!). It has been a pleasure to share so many things with them during these years, and I sincerely thank all the advice, encouragement, collaboration, and, perhaps more importantly, friendship I have received from them. Thanks so much!

I am also greatly indebted to those many others who gave me at some point important advice on my projects and scientific career, either here or abroad: Mario, Ehab, François, Wolfgang, Anne, Miguel, Franky, Kim, Axel, Bob, Marina, David, Rajee, Kyle, Enric, and many more.

I am also exceptionally grateful to all my lab mates and colleagues at the museum in Madrid: Cris, Elena, Regina, Soraya, Patrick, Fede, Gary, Fer, Íñigo, “el Buckley”, Ernesto, Lourdes, “la Pili”, and many, many more… It has been a real pleasure to work and share these years with them all. Also, I want to thank the encouragement and advice of some other friends that were always there for the good and the bad (almost always the good): Alberto, “los Davides” (Calvo, Sánchez, and Tena), Paco, Edu, Ale, Tonino, Fernando, and Hervé. Thanks guys!

Last, but not least, I really thank my parents and brothers for unconditional support and encouragement to pursue my interests and dreams, even when those dreams went beyond boundaries of field, language, and geography. This thesis, and everything else I will ever do, will always be dedicated to them.

During the course of this work, I have received financial support from the Ministry of Education and Science of Spain via the FPI predoctoral fellowship program, and the European Commission's Research Infrastructure Action via the SYNTHESYS Project.

TABLE OF CONTENTS

To Infinity… and Beyond!!

Buzz Lightyear

SUMMARY

Caecilians (order Gymnophiona) are one of the three orders or living amphibians, and probably the least known order of recent tetrapods. They are a highly specialized group with elongate, annulated bodies completely lacking limbs, and with paired sensory tentacles on each side of the snout. Most of the approximately 170 currently recognized species are tropical, soil-dwelling predators for at least their adulthood, but members of one family are secondarily adapted to aquatic habitats. Although they are a relatively small group, they have a remarkable morphological, ecological, and reproductive diversity. The caecilian fossil record is relatively poor, leaving many unresolved questions regarding the ancestry and evolution of gymnophionan lineages. There are six currently recognized families of caecilians (Rhinatrematidae, Ichthyophiidae, Uraeotyphlidae, Scolecomorphidae, Caeciliidae, and Typhlonectidae) of still controversial phylogenetic relationships. Phylogenetic inference based on morphology has been elusive because adaptation to a fossorial lifestyle makes numerous morphological characters ambiguous for many species. On the other hand, molecular phylogenetic studies on caecilians are so far limited, and mostly based on short partial sequences of ribosomal genes, so that many inferred phylogenetic relationships within caecilians are not confidently supported.

This Ph.D. thesis presents a compilation of four studies that aim to investigate the phylogeny and molecular evolution of caecilian amphibians using complete mitochondrial genomes and the nuclear *rag1* gene. The use of molecular data (particularly if large and comprehensive) has several advantages over traditional data and, despite it is not free from methodological problems and pitfalls, allows tackling phylogenetic questions long time unresolved. Furthermore, molecular systematics provides in general a powerful statistical framework for hypothesis testing and estimation of evolutionary processes.

Results from the studies presented here support that caecilians are the sister group of the Batrachia (frogs and salamanders) and suggest that the ancestors of these two lineages diverged during the Late Paleozoic, and that the caecilian crown-group originated in the Middle Mesozoic. The comprehensive mitogenomic + *rag1* dataset confidently resolve the phylogenetic relationships among the six caecilian families (although some uncertainty still remains regarding the paraphyly of Caeciliidae with respect to Scolecomorphidae). The Rhinatrematidae is the sister group of all other extant caecilians, followed by an

Ichthyophiidae + Uraeotyphlidae clade as sister group of monophyletic higher caecilians (Scolecomorphidae, Caeciliidae, and Typhlonectidae). Within these higher caecilians, Scolecomorphidae is the sister group of a paraphyletic Caeciliidae with respect to Typhlonectidae.

In general, the caecilian mitochondrial genome conforms to the vertebrate consensus mitogenomic organization, but distinct structural changes are found in the mitochondrial genomes of several species. A particular gene order rearrangement found in a group of closely related caecilians (together with comparative data for other vertebrate complete mitochondrial genomes) provides evidence that tandem duplication followed by random loss of redundant genes is the dominant mechanism of gene order rearrangement in vertebrate mitochondrial genomes, and that the genomic WANCY region is a hotspot for gene order change.

The results presented here also indicate that both complete mitochondrial genomes and the nuclear *rag1* gene are potentially useful molecular markers for the study of deep caecilian divergences.

RESUMEN (Spanish translation of Summary)

Las cecilias (orden Gymnophiona) son uno de los tres órdenes de anfibios vivos, y probablemente el orden menos conocido de tetrápodos. Se trata de un grupo altamente especializado, de cuerpo alargado y anillado, sin patas, y con tentáculos sensitivos a cada lado del hocico. La mayoría de las aproximadamente 170 especies descritas son tropicales y de hábitos subterráneos y depredadores (al menos de adultos), si bien los miembros de una familia están secundariamente adaptados a hábitos acuáticos. A pesar de ser un grupo relativamente pequeño, poseen una gran diversidad morfológica, ecológica y reproductiva. El registro fósil de las cecilias es relativamente pobre, y deja muchas cuestiones por resolver respecto al origen y evolución de los linajes de Gymnophiona. Existen seis familias actualmente reconocidas de cecilias (Rhinatrematidae, Ichthyophiidae, Uraeotyphlidae, Scolecomorphidae, Caeciliidae, y Typhlonectidae), cuyas relaciones filogenéticas son en algunos puntos muy controvertidas. La inferencia filogenética basada en morfología ha sido problemática debido a que la adaptación a hábitos subterráneos hace que numerosos caracteres morfológicos sean ambiguos para muchas especies. Por otro lado, los estudios de filogenética molecular en cecilias son hasta ahora muy limitados, y en su mayoría basados en secuencias parciales cortas de genes ribosomales, de modo que las relaciones filogenéticas inferidas dentro de las cecilias no están apoyadas con confianza.

Esta tesis doctoral presenta una recopilación de cuatro estudios que pretenden investigar la filogenia y evolución molecular en cecilias usando genomas mitocondriales completos y el gen nuclear *rag1*. El uso de datos moleculares (particularmente si son cuantiosos) tiene varias ventajas sobre los datos tradicionales y, aunque no está exento de problemas metodológicos y dificultades, permite abordar cuestiones filogenéticas no resueltas durante largo tiempo. Además, la sistemática molecular proporciona en general un marco estadístico robusto para testar hipótesis y estimar procesos evolutivos.

Los resultados de los estudios aquí presentados apoyan que las cecilias son el grupo hermano de Batrachia (ranas y salamandras) y sugieren que los ancestros de estos dos linajes divergieron durante el Paleozoico superior, y que las cecilias modernas se originaron a mediados del Mesozoico. El amplio conjunto de datos mitogenómico + *rag1* resuelve con confianza las relaciones filogenéticas entre las seis familias de cecilias (aunque queda cierta incertidumbre aún respecto a la parafilia de Caeciliidae con respecto a Scolecomorphidae).

La familia Rhinatrematidae es el grupo hermano de todas las demás cecilias modernas, seguido de un clado Ichthyophiidae + Uraeotyphlidae como grupo hermano de las cecilias superiores (que constituyen un grupo monofilético). Dentro de estas cecilias modernas, Scolecomorphidae es el grupo hermano de Caeciliidae, que aparece parafilética con respecto a Typhlonectidae.

En general, el ADN mitocondrial de las cecilias tiene la organización genómica consenso de vertebrados, pero hay características estructurales distintivas en los genomas mitocondriales de algunas especies. Una reordenación de genes particular encontrada en un grupo de cecilias estrechamente relacionadas (junto con datos comparativos para otros genomas mitocondriales completos de vertebrados) proporciona evidencia de que la duplicación en tándem seguida de pérdida aleatoria de genes es el mecanismo dominante de reordenación de genes en el genoma mitocondrial de vertebrados, y de que la región genómica WANCY es un punto con alta tasa de reordenación de genes.

Los resultados aquí presentados también indican que tanto los genomas mitocondriales completos como el gen nuclear *rag1* son marcadores moleculares potencialmente útiles para el estudio de divergencias antiguas de cecilias.

1. INTRODUCTION

1.1. *The Living Amphibians: Diversity, Origin, and Evolution*

Living amphibians (Lissamphibia) are a very successful and highly diversified group of vertebrates that includes thousands of forms (over six thousand currently recognized species; Frost, 2004) distributed throughout most terrestrial and freshwater habitats in all continents except Antarctica (Duellman and Trueb, 1994). They are divided into three orders of markedly distinct morphologies and life histories (Duellman and Trueb, 1994): the Anura (frogs and toads), the Caudata (salamanders and newts), and the Gymnophiona (caecilians) (Fig. 1). The monophyly of each of the three lissamphibian orders is widely accepted (e.g., Benton, 1990; Carroll, 1988; Duellman and Trueb, 1994; Hay et al., 1995; Trueb and Cloutier, 1991), but their origin and interrelationships are still hotly debated (Carroll, 2000; Schoch and Milner, 2004). It is clear, however, that the living amphibians possess a long evolutionary history dating back at least to the Lower Triassic, the earliest known fossils being *Triadobatrachus massinotii* from Madagascar (Rage and Rocek, 1989) and *Czatkobatrachus polonicus* from Poland (Evans and Borsuk-Bialynicka, 1998).

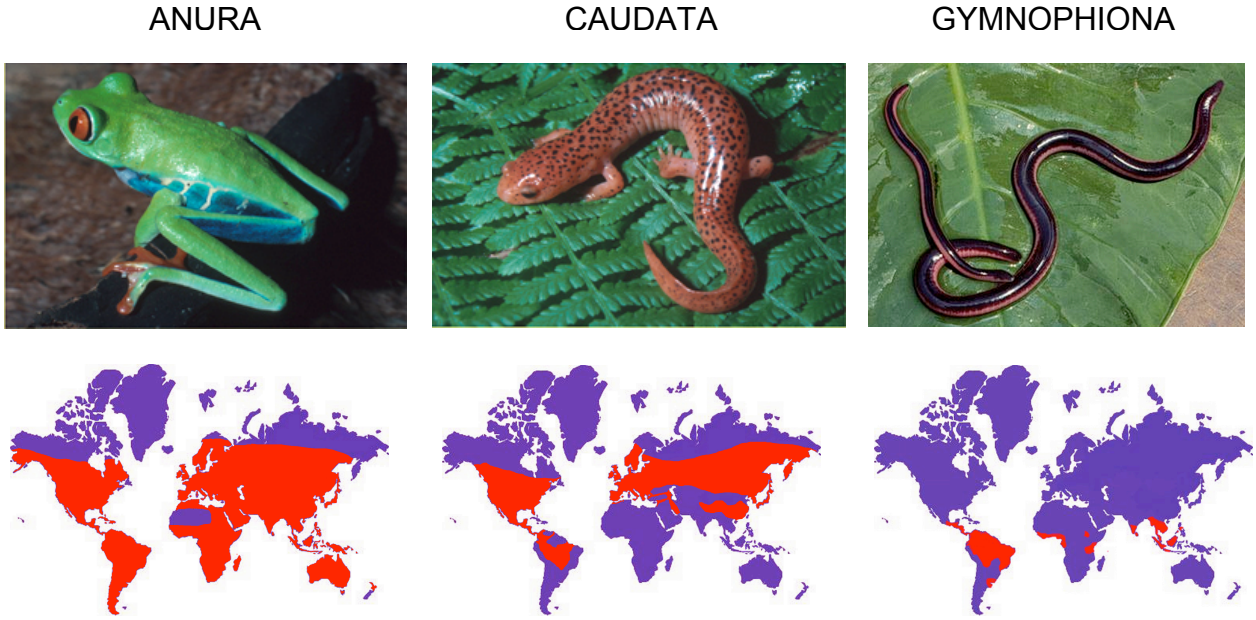


Fig. 1. The three orders of living amphibians (top), and their current distribution (in red, bottom). Species shown are: *Agalychnis callidryas* (Anura), *Pseudotriton ruber* (Caudata), *Scolecomorphus vittatus* (Gymnophiona).

The Lissamphibia are thought to constitute a monophyletic group, arisen from a single lineage of Late Paleozoic amphibians (e.g., Benton, 1990; Duellman and Trueb, 1994; Milner, 1988; Parsons and Williams, 1963; Trueb and Cloutier, 1991). Competing hypotheses have been proposed regarding the most plausible "candidate" ancestor group, either the Temnospondyli (Benton, 1990; Bolt, 1991; Milner, 1988; Panchen and Smithson, 1987; Ruta et al., 2003; Schoch and Milner, 2004; Trueb and Cloutier, 1991) or the Lepospondyli (Laurin, 1998; Laurin and Reisz, 1997), with far wider acceptance on the former (Fig. 2). Some authors have also defended a diphyletic nature of living amphibians, with frogs and salamanders arising from temnospondyl dissorophoids and caecilians from lepospondyl microsaurs (Carroll, 2001; Carroll et al., 2004) (Fig. 2).

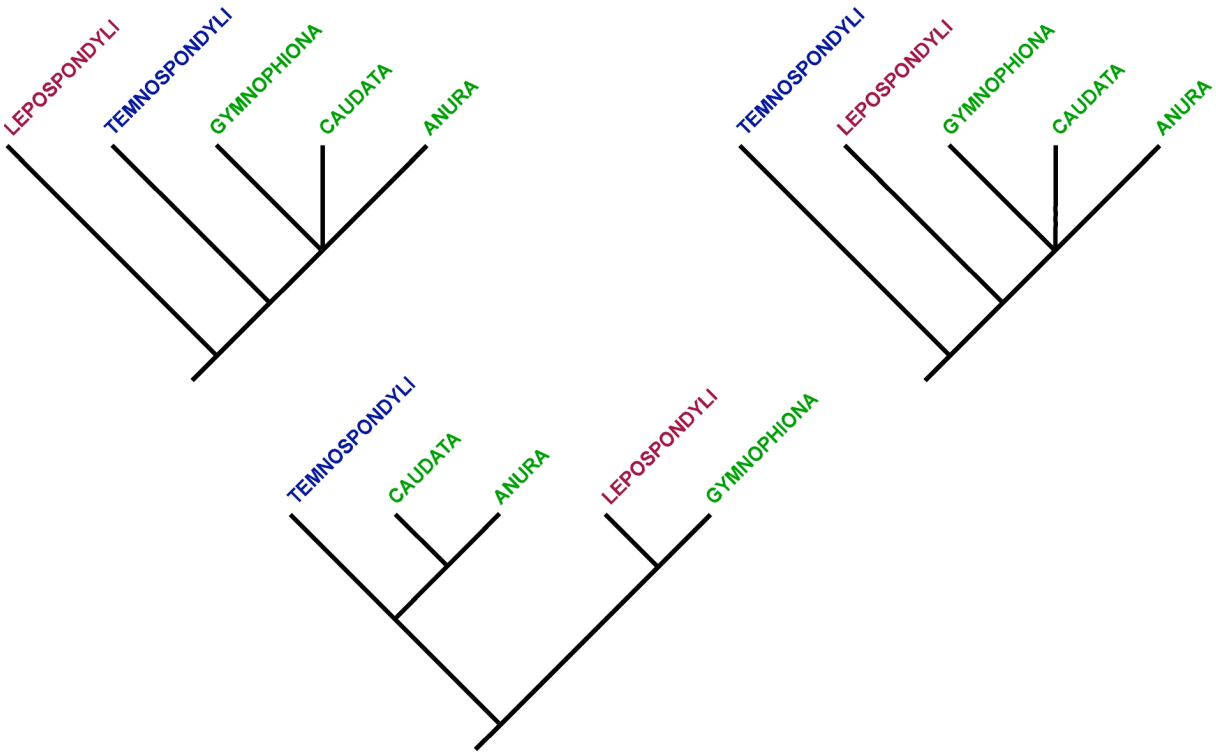


Fig. 2. Phylogenetic relationships among living and Paleozoic amphibians: Temnospondyls as ancestors of monophyletic lissamphibians (Benton, 1990; Bolt, 1991; Milner, 1988; Panchen and Smithson, 1987; Ruta et al., 2003; Schoch and Milner, 2004; Trueb and Cloutier, 1991) (top left); Lepospondyls as ancestors of monophyletic lissamphibians (Laurin, 1998; Laurin and Reisz, 1997) (top right); Lissamphibians are diphyletic, frogs and salamanders related to temnospondyls, caecilians related to lepospondyls (Carroll, 2001; Carroll et al., 2004) (bottom).

Most morphological studies (e.g., Duellman and Trueb, 1994; McGowan and Evans, 1995; Milner, 1988; Rage and Janvier, 1982; Trueb and Cloutier, 1991), and the most recent

molecular studies based on analysis of complete mitochondrial genome sequences (Zardoya and Meyer, 2001; Zhang et al., 2003), recovered frogs as the sister group of salamanders, to the exclusion of caecilians: the "Batrachia" hypothesis, (Milner, 1988) (Fig. 3). In contrast, several previous molecular studies based on (mainly partial) sequences of nuclear or mitochondrial ribosomal genes had supported a caecilian + salamander clade (Feller and Hedges, 1998; Hay et al., 1995; Hedges and Maxson, 1993; Hedges et al., 1990; Larson and Wilson, 1989): the "Procera" hypothesis (Feller and Hedges, 1998) (Fig. 3). The relatively poor fossil record of some major lissamphibian lineages (Schoch and Milner, 2004) and the fact that all three lissamphibian orders possibly acquired their specialized morphology very early in their evolutionary history (Zardoya and Meyer, 2001) have left important unresolved questions regarding the origins and evolution of living amphibians.

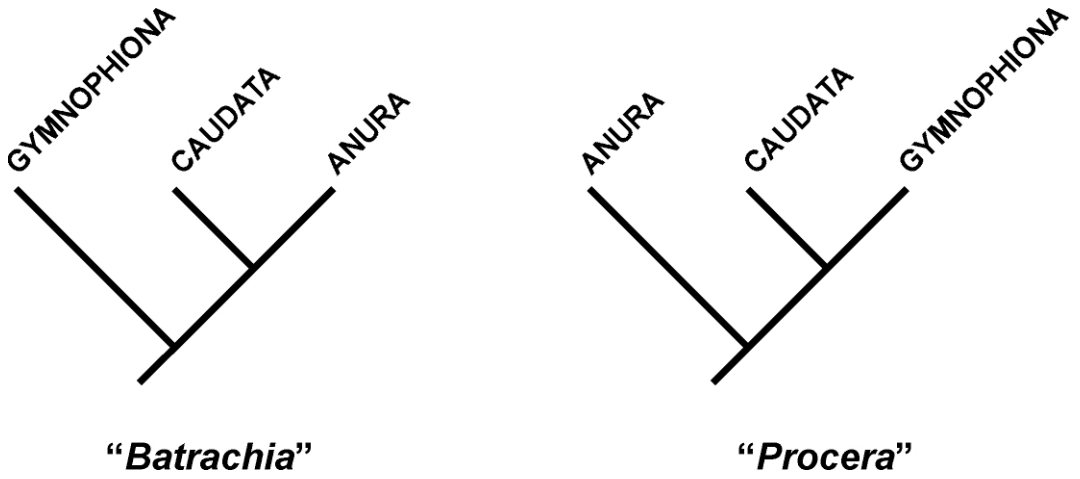


Fig. 3. Phylogenetic relationships among the three orders of living amphibians. The "Batrachia" hypothesis is supported by most morphological studies (Duellman and Trueb, 1994; McGowan and Evans, 1995; Milner, 1988; Rage and Janvier, 1982; Trueb and Cloutier, 1991) and analyses of complete mitochondrial genome sequences (Zardoya and Meyer, 2001; Zhang et al., 2003). The "Procera" hypothesis is supported by early molecular studies based on sequences of ribosomal genes (Feller and Hedges, 1998; Hay et al., 1995; Hedges and Maxson, 1993; Hedges et al., 1990; Larson and Wilson, 1989).

Frogs constitute the most diverse order of living amphibians (over five thousand currently recognized species; Frost, 2004), and possess a clear Pangaean distribution pattern (Duellman and Trueb, 1994) (Fig. 1). They were traditionally divided into the Archaeobatrachia ("primitive" frogs, containing superfamilies Leiopelmatioidea, Discoglossoidea, Pipoidea, and Pelobatoidea) and the Neobatrachia ("advanced" frogs, containing superfamilies Hyloidea and Ranoidea) (Duellman, 1975). Laurent (1979)

proposed a third suborder by distinguishing the Mesobatrachia (that included two "morphologically transitional" superfamilies, namely the Pipoidea and the Pelobatoidea) from the Archaeobatrachia. Phylogenetic reconstructions based on morphological characters generally recovered the archaeobatrachians as a paraphyletic assemblage with respect to the crown-group neobatrachians (e.g., Duellman and Trueb, 1994; Ford and Cannatella, 1993; Haas, 2003; Pugener et al., 2003). Molecular studies based on analysis of mitochondrial ribosomal genes (Hay et al., 1995; Hedges and Maxson, 1993) suggested monophyly of the Archaeobatrachia (including the Mesobatrachia), but nuclear data sets indicated paraphyly of archaeobatrachians yet again (Hillis et al., 1993; Hoegg et al., 2004). Archaeobatrachian lineages are usually viewed as remnants of an ancient and relatively fast radiation (Duellman and Trueb, 1994; Hoegg et al., 2004). In contrast to the Archaeobatrachia, there is compelling evidence that the Neobatrachia is a monophyletic group (Biju and Bossuyt, 2003; Duellman and Trueb, 1994; Ford and Cannatella, 1993; Hay et al., 1995; Hoegg et al., 2004; Vences et al., 2003). Within this clade, the two major superfamilies distinguished, the Hyloidea and the Ranoidea, are each species-rich radiations, very diverse morphologically and ecologically, containing many additional families (Hoegg et al., 2004). Recent molecular studies (Biju and Bossuyt, 2003; Hoegg et al., 2004) recovered both lineages as clearly monophyletic, but identified several controversial families (Heleophrynidae, Sooglosidae, Myobatrachidae, and Nasikabatrachidae) as basal early splits in the Neobatrachia that could not be confidently assigned to either Hyloidea or Ranoidea. Although the Ranoidea and the Hyloidea are nowadays widely distributed, Feller and Hedges (1998) suggested that their origin was associated with the split of Africa and South America during the Cretaceous (Pitman III et al., 1993). It is widely accepted that the high adaptative success of frog radiations, not remotely paralleled by caecilians or salamanders, is related to the early acquisition of two major key innovations (already present in the most ancient lineages): powerful hind limbs and short stiffened vertebral column (the urostyle) for jumping locomotion, and a larval form that differs drastically from the adult in morphology and ecology (thus avoiding niche competence) (Duellman and Trueb, 1994).

Salamanders are a less diversified group of living amphibians than frogs (approaching six hundred currently recognized species; Frost, 2004), and are mostly distributed on Laurasian-derived landmasses (Duellman and Trueb, 1994) (Fig. 1). Despite being regarded as the best known lissamphibian group, their phylogeny remains fairly problematic, particularly due to confusing effects of paedomorphy on character interpretation. It seems paedomorphy has independently evolved in representatives of at least five families of

salamanders, namely Cryptobranchidae, Sirenidae, Proteidae, Ambystomatidae, and Salamandridae (Duellman and Trueb, 1994). Most studies (both morphological and molecular) have traditionally recovered three families (Cryptobranchidae, Hynobiidae, and Sirenidae) at the base of the salamander phylogenetic tree (Duellman and Trueb, 1994; Gao and Shubin, 2001; Hay et al., 1995; Hedges and Maxson, 1993; Larson and Dimmick, 1993), with Cryptobranchidae and Hynobiidae confidently recovered as sister taxa to each other. These three lineages have all external fertilization and angular and prearticular bones of the lower jaw not fused, which are considered to be "primitive" traits (Duellman and Trueb, 1994). The "basal" condition of Sirenidae and Cryptobranchidae indicates that paedomorphic lifestyle (almost exclusive of salamanders among amphibians) was already present in the early evolutionary history of salamanders. Only early molecular studies (Larson, 1991; Larson and Wilson, 1989) recovered two other families (Amphiumidae and Plethodontidae) at the base of the salamander tree. Despite the phylogeny of salamander lineages still remains controversial (particularly the position of several lineages, such as Amphiumidae and Proteidae), it seems fairly clear that all salamander families arose in the Laurasian part of Pangaea (Duellman and Trueb, 1994). Feller and Hedges (1998) suggested that salamanders arose in the Mesozoic by vicariance directly linked to the breakup of supercontinent Pangaea, but both the fossil record and molecular evidence (Benton, 1990; Carroll, 2001; Duellman and Trueb, 1994; Evans et al., 1996; Milner, 1988; Zardoya and Meyer, 2001) points at a much earlier origin of salamanders, predating the Mesozoic continental fragmentation. Salamanders, with their slender bodies, well developed tails, and proportionally paired limbs, have likely retained the most similar morphology to the ancestral tetrapod body form (Duellman and Trueb, 1994).

1.2. *The Caecilians: Diversity and Phylogenetic Controversies*

Caecilians (order Gymnophiona) are the least known order of living amphibians, and probably of recent tetrapods. They are readily distinguished from frogs and salamanders by their elongate, annulated bodies completely lacking limbs, and protusible sensory tentacles on each side of the snout (Duellman and Trueb, 1994; Himstedt, 1996; Noble, 1931; Taylor, 1968) (Fig. 4). Most of the approximately 170 currently recognized species (Frost, 2004; Nussbaum and Wilkinson, 1989) are tropical, soil-dwelling predators for at least their adulthood, but members of the South American family Typhlonectidae are semiaquatic or

aquatic (Taylor, 1968; Wilkinson and Nussbaum, 1999). Most caecilian species possess a heavily ossified, roofed skull (Fig. 4), and some have reduced eyes (covered by skin or bone), which has been interpreted by some authors as adaptation for burrowing (Duellman and Trueb, 1994; Nussbaum, 1998; Nussbaum and Wilkinson, 1989). Although they are a relatively small group, it is clear that they have a remarkable morphological (Taylor, 1968; Wilkinson and Nussbaum, 1997), ecological (Gower et al., 2004; Loader et al., 2003), and particularly reproductive (Kupfer et al., 2004; Wake, 1977) diversity (caecilians are known to present both oviparity, with an aquatic larva or direct development, and viviparity; Fig. 4).

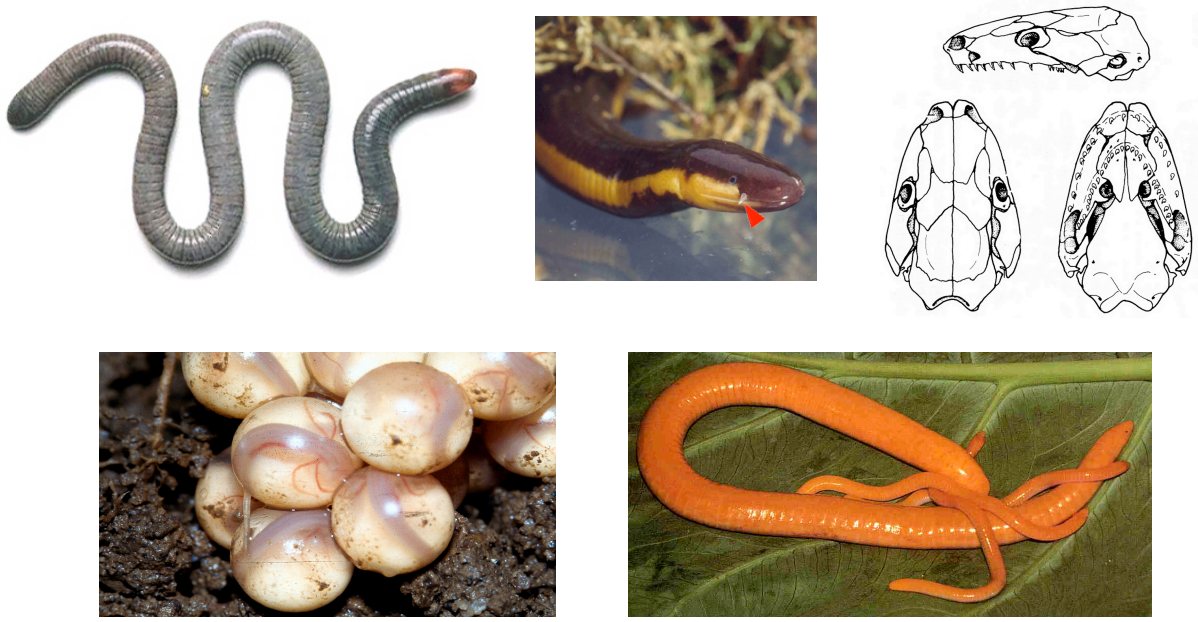


Fig. 4. Caecilian diversity: adult of *Herpele squalostoma* (top left); head of *Ichthyophis* cf. *kohtaoensis* (top centre), red arrow indicates sensory tentacle; skull of *Uraeotyphlus* sp. (top right); eggs of oviparous *Gegeneophis ramaswamii* (bottom left); adult female of viviparous *Schistometopum thomense* with offspring (bottom right).

Fossils of potential crown-group caecilians are represented only by isolated vertebrae of uncertain affinities from the Paleocene of Brazil and Bolivia (Estes and Wake, 1972; Rage, 1986; Rage, 1991), and the Upper Cretaceous of Sudan (Evans et al., 1996; Werner, 1994). In contrast, more complete specimens of putative stem-group caecilians have been found: *Eocaecilia micropodia* from the Lower Jurassic of Arizona (Jenkins and Walsh, 1993), and *Rubricacaecilia monbaroni* from the Lower Cretaceous of Morocco (Evans and Sigogneau-Russel, 2001). These two fossil putative stem-group caecilians were elongate and, interestingly, retained limbs (at least *Eocaecilia*; Fig. 5), which suggest that limblessness in

caecilian ancestors evolved subsequent to body elongation. Albeit the great paleontological value of *Eocaecilia* and *Rubricacaecilia*, the scarce fossil record (particularly of crown-group caecilians) leaves many unresolved questions regarding the origin, evolution, and historical biogeography of gymnophionan lineages.

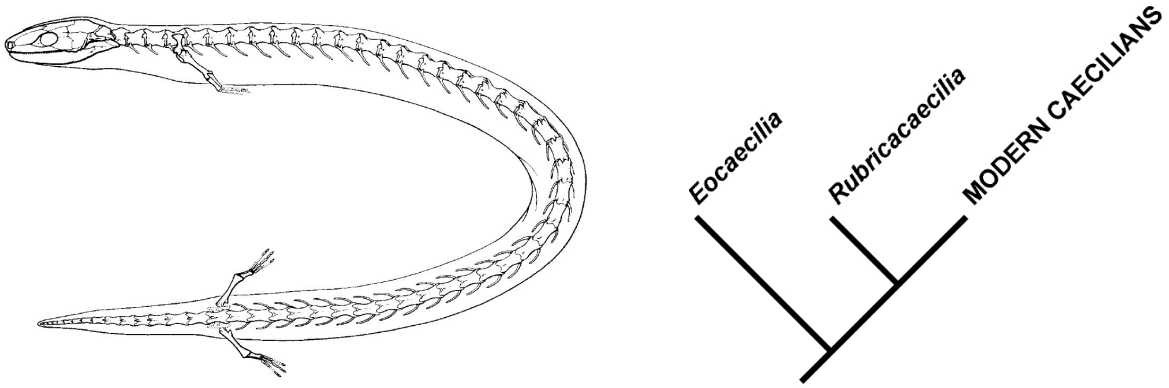


Fig. 5. Reconstruction of *Eocaecilia micropodia* (left) according to Jenkins and Walsh (1993), and postulated phylogenetic relationships among putative stem-group and modern caecilians (right) according to Evans and Sigogneau-Russel (2001).

For many years, all living caecilians were placed into a single family, namely Caeciliidae, until the seminal monograph of Taylor (1968). This thorough work prompted new comparative and phylogenetic studies based on morphological and life history data (Duellman and Trueb, 1994; Nussbaum, 1977; Nussbaum, 1979; Nussbaum and Wilkinson, 1989; Taylor, 1969; Wilkinson and Nussbaum, 1999) and a complete revision of caecilian systematics. The single family Caeciliidae was partitioned into six higher taxa that are mostly recognized as families today (Duellman and Trueb, 1994; Nussbaum and Wilkinson, 1989). Four caecilian families have relatively restricted distributions (Fig. 6): the South American Rhinatrematidae (two genera, nine species) and Typhlonectidae (five genera, 13 species), Indian Uraeotyphlidae (one genus, five species), and African Scolecomorphidae (two genera, six species). Ichthyophiidae (two genera, 40+ species) occur in South and South East Asia (East of Wallace's line; Fig. 6). The more cosmopolitan Caeciliidae (21 genera, 100+ species) occur on all landmasses where caecilians occur except South East Asia (Fig. 6). The four smaller and more local families comprise morphologically distinctive caecilian clades (Nussbaum, 1977; Nussbaum, 1979; Nussbaum, 1985; Wilkinson and Nussbaum, 1999). In contrast, molecular data suggest that Ichthyophiidae might not be monophyletic (Gower et al.,

2002), and morphology and molecules agree that the Caeciliidae, which comprises those caecilians that have not been assigned to the five more recently described families, is paraphyletic with respect to Typhlonectidae (Hedges et al., 1993; Nussbaum, 1979; Wilkinson, 1997; Wilkinson et al., 2002; Wilkinson et al., 2003).

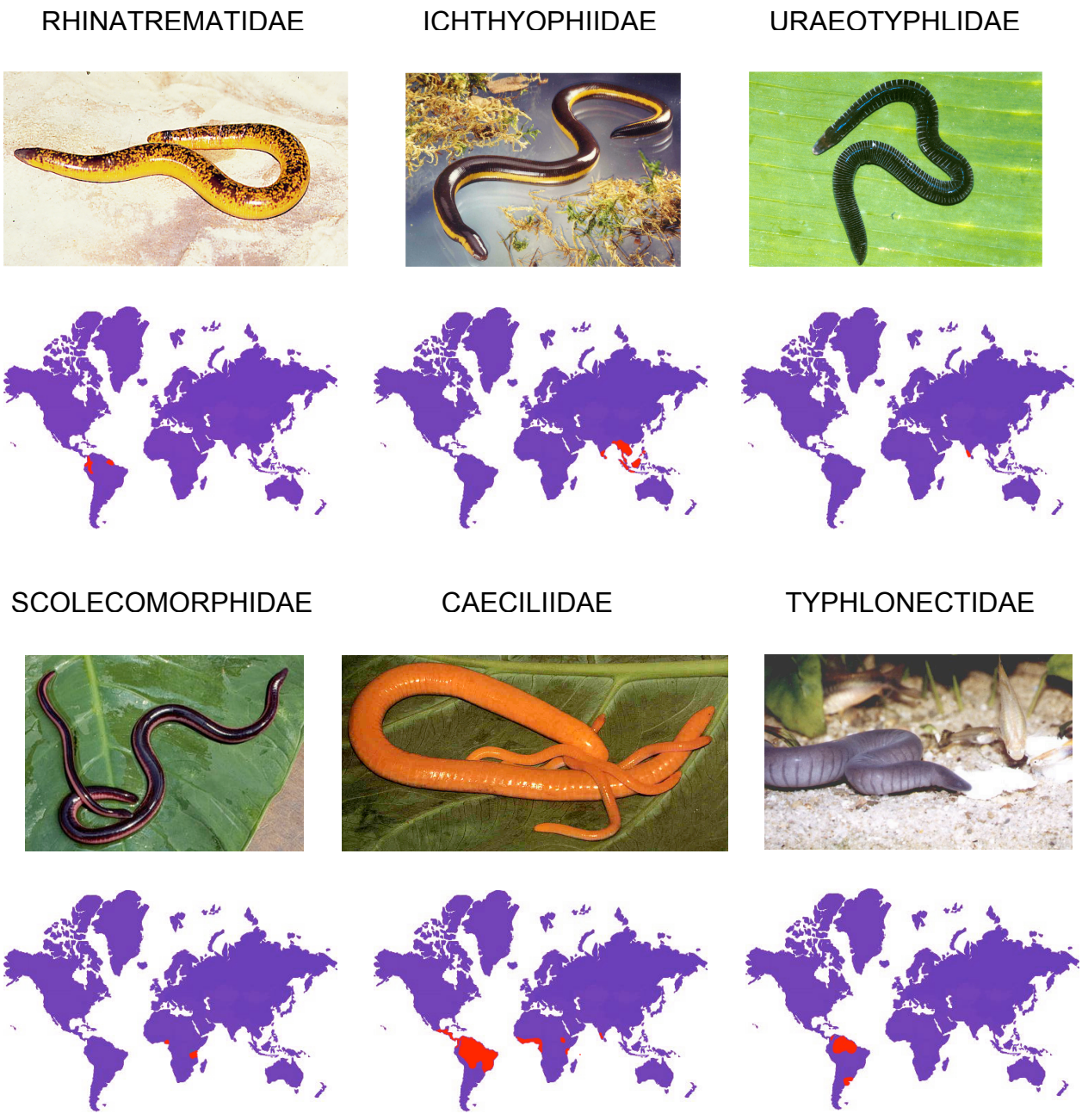


Fig. 6. The six currently recognized families of modern caecilians. Present distribution is indicated in red. Species shown are: *Epicrionops* sp. (Rhinatrematidae), *Ichthyophis* cf. *kohtaoensis* (Ichthyophiidae), *Uraeotyphlus* sp. (Uraeotyphlidae), *Scolecomorphus vittatus* (Scolecomorphidae), *Schistometopum thomense* (Caeciliidae), *Typhlonectes natans* (Typhlonectidae).

The current distribution of extant caecilians (Fig. 1) is consistent with a Gondwanan origin of the Order (Duellman and Trueb, 1994; Hedges et al., 1993; Wilkinson et al., 2002). However, the discovery of *Eocaecilia* in North America offers the possibility of a Pangaean distribution of (at least stem-group) caecilians by the end of the Triassic (Jenkins and Walsh, 1993). On present evidence, an origin in the Gondwanan part of Pangaea with subsequent northwards dispersal and an origin in the Laurasian part with southwards dispersal are both plausible (Evans and Sigogneau-Russel, 2001), and more paleontological data is needed to confidently rule out one of these two hypotheses. According to Duellman and Trueb (1994) and Hedges et al. (1993), the presence of caeciliids in South America, Africa, Seychelles, and India suggests that the split of the major caecilian lineages occurred prior to the breakup of Gondwana. A successive dispersal from the Indian Plate subsequent to its collision with Asia has been proposed to explain the origin of ichthyophiid caecilians in South East Asia (Duellman and Trueb, 1994; Gower et al., 2002; Wilkinson et al., 2002). Similarly, a dispersal from South America has been proposed to explain the distribution of some caeciliids in Central America (Duellman and Trueb, 1994).

Nussbaum (1979) presented the first numerical phylogenetic analysis of caecilians, using morphological characters to investigate the interrelationships of 12 genera. This, and the subsequent analyses of Duellman and Trueb (1994) and Hillis (1991) that used family level taxa and a subset of Nussbaum (1979) characters, identified a clade comprising the caeciliids, typhlonectids and scolecomorphids that Nussbaum (1991) dubbed the "higher" caecilians (Fig. 7). The Uraeotyphlidae, Ichthyophiidae and Rhinatrematidae were successively more distant outgroups to the higher caecilians in these analyses. Diverse morphological evidence that the Rhinatrematidae is the sister group of all other extant caecilians (Nussbaum, 1977; Wilkinson, 1992; Wilkinson, 1996) is considered to provide strong support for this hypothesis (Fig. 7), which has been used to root caecilian phylogenetic trees in more recent morphological and molecular analyses (Gower et al., 2002; Wilkinson, 1997; Wilkinson et al., 2002; Wilkinson et al., 2003; Wilkinson and Nussbaum, 1996). Wilkinson and Nussbaum (1996) and Wilkinson (1997) also supported the monophyly of the higher caecilians, but found strong support for an alternative arrangement of more deep-branching families, in which the Ichthyophiidae and Uraeotyphlidae are sister taxa (in contrast to the arrangement recovered by Duellman and Trueb (1994) (Fig. 7). Whereas earlier family-level phylogenies (Duellman and Trueb, 1994; Hillis, 1991) recovered Caeciliidae and Typhlonectidae as more closely related to each other than to Scolecomorphidae, the most comprehensive morphological study to date (Wilkinson, 1997)

was unable to confidently resolve phylogenetic relationships among these higher caecilians (Fig. 7).

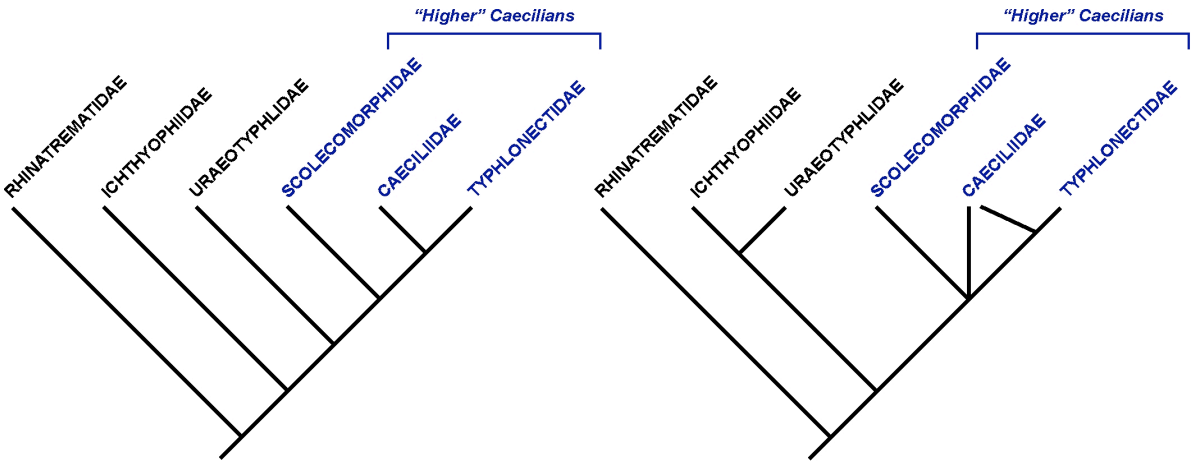


Fig. 7. Hypotheses for the phylogenetic relationships of caecilian families: (left) relationships proposed on the basis of early cladistic analyses of morphological characters (Duellman and Trueb, 1994; Hillis, 1991; Nussbaum, 1979); (right) alternative hypothesis based on more recent morphological (Wilkinson, 1997; Wilkinson and Nussbaum, 1996) and molecular (Wilkinson et al., 2002; Wilkinson et al., 2003) analyses.

Previous molecular analyses that have been informative regarding the relationships among caecilian families have used partial nucleotide sequences of mitochondrial ribosomal genes (Hay et al., 1995; Hedges and Maxson, 1993; Hedges et al., 1993; Wilkinson et al., 2003; Wilkinson et al., 2002). Only a single study also included partial sequences of mitochondrial *cob* (Gower et al., 2002). These studies have supported recent morphological analyses in recovering a monophyletic Ichthyophiidae + Uraeotyphlidae and Nussbaum's (1991) higher caecilians (caeciliids, scolecomorphids and typhlonectids) (Wilkinson et al., 2003), and a paraphyletic Caeciliidae (Hedges and Maxson, 1993; Hedges et al., 1993; Wilkinson et al., 2002; Wilkinson et al., 2003) (Fig. 7). Wilkinson et al. (2003) carried out the only previous molecular analysis to include members of all six currently recognized families. In agreement with the most recent morphological investigation, their study suggested that Caeciliidae may be paraphyletic with respect to Scolecomorphidae as well as Typhlonectidae. However, many relationships within the higher caecilians were not strongly supported, and they suggested that more molecular and morphological data were required to resolve these relationships.

1.3. *Molecular Systematics*

In recent years, the outstanding advancement of molecular biology and bioinformatics has supplied systematic biologists with a bunch of powerful tools for tackling phylogenetic problems long time unresolved. Molecular systematics could be defined as the use of information contained in molecular data to reconstruct phylogenetic relationships (Page and Holmes, 1998). Of the various patterns and processes of molecular evolution that can be used for molecular systematics (reviewed by Hillis et al., 1996; Rokas and Holland, 2000), the analysis of deoxyribonucleic acid (DNA) and/or protein sequence variation has increasingly become a standard in the vast majority of recent phylogenetic studies.

The use of DNA and amino acid sequences in molecular systematics has several advantages over traditional morphological approaches (Hillis and Wiens, 2000): universality of character types and states (more objective selection and definition of homologous sets), high number of characters available for analyses (better statistical performance of the data), high degree of substitution rates among genes and among gene regions (thus providing different levels of variability for specific questions), increasingly comprehensive knowledge of the molecular basis underlying sequence evolution and function (thus allowing the construction of more efficient models of the evolutionary process), relative easy collection of the data in different taxa (even from very small tissue samples) and by researchers that do not necessarily have taxon-specific expertise. Furthermore, the vast amount of sequence information that is being generated in recent years for virtually every taxon has definitely boosted the possibilities for comparative and phylogenetic studies.

However, phylogenetic inference from molecular data is not free from methodological problems and pitfalls (reviewed by Hillis and Wiens, 2000; Maley and Marshall, 1998; Stevens and Schofield, 2003). One of the major criticisms to molecular systematics comes from the fact that molecular sequence data possess a relatively low character state space (four states in the case of DNA, 20 in the case of amino acid), which may entail a high probability of homoplasy due to saturation of the substitution process (e.g., at a given position, two sequences might have the same character state just by chance, and not due to common ancestry). Another important issue is that gene phylogenies do not necessarily match those of the organisms. This will only be true if the set of genes used to reconstruct the phylogeny are orthologous (i.e. homologous genes that have evolved independently from a common ancestor that did not undergo a gene duplication). Genes that are homologous because of duplication are termed paralogous. Additional criticisms to molecular systematics are that molecular data

cannot be obtained for ancient fossil taxa (although some data have been obtained for "recently" extinct organisms; Pääbo et al., 2004), and that they cannot be used on their own for describing new species.

Both molecular and morphological data are useful and necessary in systematics. They constitute independent and complementary sources of information for cross-validating hypotheses about evolutionary patterns and processes at different levels of biological organization.

1.3.1. *Modelling Sequence Evolution*

A model of sequence evolution provides a statistical description of the process of character state change, i.e. the process of nucleotide or amino acid substitution. In general, nucleotide and amino acid substitution is viewed as a Markov process: a mathematical model of infrequent changes of discrete states over time, in which future events occur by chance and depend only on the current state, and not on the history of how the state was reached (Felsenstein, 1981; Swofford et al., 1996; Whelan et al., 2001). This Markov model also assumes that substitution rates do not change over time (time-homogeneous), and that relative frequencies of each character state are at equilibrium (stationarity) (Posada, 2003). The mathematical expression of a substitution model is a table of rates (substitutions per site per unit of evolutionary distance) at which each character state (either of nucleotide or amino acid) is replaced by each alternative state (Swofford et al., 1996). As models become more sophisticated, these instantaneous rate matrices grow in complexity, and other parameters can be incorporated, such as frequency parameters (Yang, 1994a), and among-site rate variation parameters (Reeves, 1992; Yang, 1994b).

Over the years, an array of models of increasing complexity has been described, both for nucleotide and amino acid sequences (see Felsenstein, 2004). In general, two main approaches have been followed to building models of sequence evolution (Whelan et al., 2001): empirically using properties calculated through comparison of large numbers of observed sequences, or parametrically on the basis of the chemical or biological properties of DNA or amino acids. Empirical models result in fixed parameter values that are estimated only once and then assumed to be applicable to all datasets. Modelling of amino acid replacement, such as mtREV (Adachi and Hasegawa, 1996) and JTT (Jones et al., 1992), has concentrated on this empirical approach. In contrast, parametric models allow the parameter models to be derived from the dataset in each particular analysis. Modelling of nucleotide

replacement, such as HKY (Hasegawa et al., 1985) and GTR (Rodríguez et al., 1990), has concentrated on this parametric approach. More complex models of sequence evolution (such as codon-based models, or models attempting to accommodate structural elements of the analyzed molecules) have also been described, and many more are to come.

A proper characterization of the process of sequence evolution is essential in molecular phylogenetic inference (Cunningham et al., 1998), as phylogenetic methods tend to be less accurate or inconsistent when a wrong model of sequence evolution is assumed (Bruno and Halpern, 1999; Huelsenbeck and Hillis, 1993). In general, model selection strategies attempt to find the appropriate level of complexity on the basis of the available data (Holder and Lewis, 2003). Increasing model complexity improves the fit to the data, but also increases variance in estimated parameters (Huelsenbeck and Crandall, 1997; Posada and Crandall, 2001). In recent years, several statistical methods (based on hypotheses testing) have been developed for selecting best-fit models of sequence evolution for a given dataset (Posada, 2003; Posada and Crandall, 2001). These methods use likelihood ratio tests (see below) or information criteria (such as the Akaike information criterion, AIC; Akaike, 1973) to contrast the fit to the data of different alternative models.

1.3.2. *Methods of Phylogenetic Inference*

Given a particular group of organisms, the process of phylogenetic estimation starts with the collection of homologous sequence data (both new and, if available, previously determined one – usually downloaded from the many available gene databases). Typically, a few outgroup sequences are included to root the tree (indicating which nodes in the tree are the oldest, and providing clues about ancestral sequence states). The next step is to align the sequences (i.e. adding gaps to a matrix) so that the characters (either nucleotides or amino acid) at each position (column of the matrix) are related to each other by descent from a common ancestral residue (positional homology). This step is critical as the rest of the phylogenetic inference process relies on it (Goldman, 1998; Phillips et al., 2000) and a bunch of algorithms have been developed for multiple-sequence alignment (reviewed by Thompson et al., 1999). In addition to the sequence data, a model of sequence evolution must be chosen, as the methods used in molecular phylogenetics are based on a series of assumptions about how the substitution process works (see above). These assumptions can be implicit, like in parsimony methods, or explicit, like in distance and likelihood-based methods (Posada and Crandall, 2001).

Among the various methods developed to reconstruct phylogenetic relationships (see reviews by Felsenstein, 2004; Holder and Lewis, 2003; Swofford et al., 1996; Whelan et al., 2001), there are four that have largely dominated molecular systematic studies: maximum parsimony, neighbour-joining, maximum likelihood, and Bayesian inference.

- **Maximum parsimony** (Farris, 1983; Fitch, 1971): This is one of the earliest inference methods that comes directly from Hennig's (1950) cladistic analysis. Maximum parsimony uses directly character states without conversions (unlike distance methods – see below), and is based on an optimality criterion (a rule to decide which of two trees is the best): it selects the tree or trees requiring the fewest character state changes (thus attempting to minimize homoplasy). In this method, the tree space is usually searched using heuristic searches or, when the number of sequences is small (<12), exhaustive searches (Felsenstein, 2004). The advantage of this method is that it is fast enough for the analysis of large datasets containing many sequences, and it is robust if branches of the tree are short (either because sequences are closely related or because the taxon sampling is dense). However, it can perform poorly (even seriously misleading) if there is substantial variation in rates of evolution among taxa. In this case, taxa with the fastest substitution rates appear in the tree as long branches, and tend to artefactually attract one another. This phenomenon is called long-branch attraction (LBA; Felsenstein, 1978), and parsimony is particularly affected by it (Huelsenbeck, 1997; Swofford et al., 1996). Unweighted parsimony lacks an explicit model of sequence evolution (Goldman, 1990), thus it is difficult for this method to deal with high degree of homoplasy (i.e. parallel, convergent, reversed, or superimposed changes) when markedly divergent sequences are analysed. In such a case, parsimony analyses can be weighted to incorporate (through the use of step matrices) prior assumptions about the costs of character state change (Swofford et al., 1996).

- **Neighbour-joining** (Saitou and Nei, 1987): This is a pairwise distance method based on the assumption that dissimilarity between two sequences (evolutionary distance, i.e. the number of changes that have occurred along the branches) is directly related to their phylogenetic relationship. In this method, the DNA or amino acid sequences are first converted into a distance matrix that is then used to reconstruct a phylogenetic tree. Neighbour-joining is a clustering method rather than an optimality method, and hence it does not optimise the fit between the inferred tree and the data. The main advantage of this method is that it is relatively fast (compared to all other methods available), and performs well when the

divergence between sequences is low. Disadvantages of distance methods include the loss of information when the sequences are converted to distances, and the difficulty in obtaining reliable estimates of pairwise distances for highly divergent sequences. Neighbour-joining is a particular case of another distance method based on optimality, **minimum evolution** (Rzhetsky and Nei, 1992), which employ heuristic searches to find the tree with the smallest sum of branch lengths (the minimum evolution score). Both neighbour-joining and minimum evolution can incorporate models of evolution to correct pairwise genetic distances for multiple substitutions at the same site (Felsenstein, 2004; Nei and Kumar, 2000).

- **Maximum likelihood** (Felsenstein, 1981): This method is one of the standard tools of statistics. In the context of molecular systematics, the likelihood of a phylogenetic tree is the probability of observing the data (set of sequences being analyzed) given the tree and the model of evolution. This is an optimality method too: the tree that renders the observed sequences the most likely under the assumed evolutionary model is better. As for parsimony, the tree space is usually explored using heuristic searches. The great advantage of maximum likelihood is that it allows the inference of phylogenetic trees using complex models of sequence evolution (including the ability to estimate model parameters, hence allowing simultaneous inference of patterns and processes of molecular evolution), and provides a powerful statistical framework for hypotheses testing (see below). The strong statistical foundations of likelihood-based methods probably makes them the most robust way for estimating molecular phylogenies and understanding sequence evolution (Whelan et al., 2001). However, there are also criticisms to this method related to the fact that it can be prohibitively slow and computationally demanding, and that the result is especially dependent on the correctness of the employed model of sequence evolution (Holder and Lewis, 2003; Whelan et al., 2001). Because likelihood values are often very small, they are usually expressed as log likelihoods, ln L (computationally easier to handle).

- **Bayesian inference** (Huelsenbeck et al., 2001; Rannala and Yang, 1996): This is the most recent of all phylogenetic inference methods. The field of Bayesian statistics is closely allied with maximum likelihood: the optimal hypothesis is the one that maximizes the posterior probability. According to Bayes' theorem, the posterior probability for a hypothesis is proportional to the likelihood multiplied by the prior probability of that hypothesis. Bayesian analysis allows complex models of sequence evolution to be implemented for the whole sequence dataset, and for different partitions of it. This method involves specifying a model

and a prior distribution and then integrating the product of these quantities over all possible parameter values to determine the posterior probability for each tree. However, the likelihood functions for phylogenetic models are currently too complex to integrate analytically, so Bayesian approaches rely on Markov chain Monte Carlo (MCMC) procedures. This algorithm works by taking a series of steps of a conceptual chain for approximating probability distributions on a wide variety of contexts. Bayesian inference has the advantage of a strong connection with the likelihood framework and its powerful statistical foundations, but being faster and computationally less requiring using equally (or even more) complex models of sequence evolution. Moreover, as a result of the MCMC process, there is a posterior probability associated to each node on the inferred Bayesian tree (the fraction of times a clade occurs among the sampled trees) that can be used as a measure of support for that node. Disadvantages of Bayesian methods come from the fact that prior distributions for parameters must be specified, and that it can be difficult to determine whether the MCMC approximation has run for a sufficient number of cycles (Holder and Lewis, 2003). Some criticisms to Bayesian inference are also related to the putative overconfidence of posterior probability measures of node support (Suzuki et al., 2002), and recommendations have been done that posterior probabilities should only be considered reliable if greater than 0.95 (Alfaro et al., 2003; Erixon et al., 2003; Huelsenbeck and Rannala, 2004).

Not counting Bayesian inference (that yields a tree with support values for each node, measured as posterior probabilities), all other methods of phylogenetic reconstruction produce only point estimates of the phylogeny. However, an important issue is to know how strongly the data support each of the relationships depicted in the tree. Several methods for assessing confidence exist (Goldman et al., 2000), but this issue has been traditionally tackled by bootstrapping (first applied to phylogenetics by Felsenstein, 1985). This is a statistical resampling technique by which distributions that are difficult to calculate exactly can be estimated by the repeated creation and analysis of artificial datasets. To assess node support in phylogenetics, non-parametric bootstrapping is used: new datasets are created by sampling randomly and with replacement from the original data (these new bootstrap datasets are of the same size as the original); a desired quantity of bootstrap datasets is computed (typically between 500 and 2000; Hedges, 1992; Zharkikh and Li, 1992) and the resulting distribution is used to estimate the dispersion that would be expected if the same number of new independent datasets had been collected. The exact interpretation of the statistical significance of bootstrap proportions is elusive, but several authors (Hillis and Bull, 1993;

Zharkikh and Li, 1992) have proposed that they are conservative measures of support, so a value of 70% or greater might indicate substantial confidence for a group.

1.3.3. *Hypothesis Testing in Phylogenetics*

One of the most appealing topics in molecular systematics is the availability of methods for the statistical testing of competing phylogenetic hypotheses. These methods are available almost exclusively within the likelihood framework, although some tests have also been developed for other frameworks, such as parsimony (Templeton, 1983). They allow assessment of which model provides the best fit for a given dataset, and the degree of confidence we have in any given topology being the true topology.

One of the methods to compare two competing hypothesis is the likelihood ratio test (LRT; Felsenstein, 1981; Huelsenbeck and Crandall, 1997), which has been extensively used for selecting competing best-fit models of sequence evolution for a given dataset, and for testing deviations from clock-like evolution (global molecular clock hypothesis – see below). Competing hypotheses are compared using a statistic, 2Δ (calculated as the ratio of the likelihood scores of the alternative hypothesis to the null hypothesis), that measures how much better an explanation of the data the alternative hypothesis gives. In order to perform a significance test, the distribution of 2Δ values expected under the simpler hypothesis is required. If the two competing hypotheses are nested (that is, the null hypothesis is a special case of the alternative hypothesis), then the 2Δ distribution is asymptotically distributed as a χ^2 (or, in some cases, a mixed χ^2; Goldman and Whelan, 2000; Ota et al., 2000) with the number of degrees of freedom equal to the difference in the number of parameters between the two models.

When the hypotheses being compared are not nested, the χ^2 approximation may perform poorly. In this case, the null distribution of the LRT statistic can be approximated by parametric bootstrapping (Efron, 1985; Goldman, 1993; Huelsenbeck et al., 1996). Unlike non-parametric bootstrap (where datasets are generated by resampling from the original data), the parametric bootstrap uses Monte Carlo simulation to generate the data. Replicate datasets of the same size as the original (usually 200-1000) are simulated according to the null hypothesis being tested. For each replicate dataset, the likelihoods according to both the null and alternative hypotheses are estimated, and the LRT statistic is calculated. These simulated 2Δ values form the null distribution of the LRT statistic, allowing implementation of a

significance test. The main disadvantage of parametric bootstrapping is that it is computationally demanding, and even unfeasible when large datasets are considered.

Apart from parametric bootstrapping, there have been developed several non-parametric likelihood-based tests to determine whether the difference in fit of two or more alternative tree topologies (always non nested hypotheses) to the data is significantly greater than expected under the null hypothesis of random sampling error. Of the various methods of this kind, the most widely used are the Kishino-Hasegawa (Kishino and Hasegawa, 1989), the Shimodaira-Hasegawa (Shimodaira and Hasegawa, 1999), and the approximately unbiased (Shimodaira, 2002) tests. They all are based on the estimation of LRT statistics, and use different non-parametric bootstrapping procedures (resampling at one or multiple scales) for assessing their variance, and obtaining an estimation of their distribution (thus permitting significance evaluation). Empirical comparisons of non-parametric and parametric bootstrapping tests appear to indicate that the former tend to be conservative (i.e. unwilling to reject topologies as untrue) because of multiple comparisons and deviations from some of their basic assumptions, and that the latter tend to be liberal (i.e. willing to reject topologies as untrue) because of the use of oversimplified models of sequence evolution to construct the null distribution (Buckley, 2002; Goldman et al., 2000; Strimmer and Rambaut, 2001).

1.3.4. *Estimation of Divergence Times*

A key feature of molecular phylogenies is that not only relationships can be reconstructed, but also that divergence events can be dated using various models of the expected rate of accumulation of mutations in the sequences over time. The idea of dating evolutionary divergences using calibrated sequence distances was first proposed by Zuckerkandl and Pauling (1965) who postulated that the amount of difference between the DNA molecules of two species is a function of the time since their evolutionary separation. This was termed “molecular clock” and was shown comparing amino acid substitution rates with ages estimated from fossils. The central assumption of the molecular clock is that all branches of a phylogenetic tree evolve at the same, global substitution rate (i.e. there is rate constancy). A clock-like tree is ultrametric (i.e. the total distance between the root and every tip is constant), so nodal depths can be easily dated if the divergence time for at least one node is known (calibration point): the global rate of substitution is calculated and, based on it, divergence times for all nodes can be estimated by linear regression of the molecular distances (Li and Graur, 1991; Nei, 1987). If several calibration points are used, then a

regression line (whose slope is a function of the global substitution rate) is built, and the divergence times for the unknown nodes are interpolated (or extrapolated). The molecular clock hypothesis is in perfect agreement with the neutral theory of evolution (that postulates that the majority of substitutions in genes are the result of random fixation of selectively neutral mutations; Kimura, 1968; Kimura, 1983).

Unfortunately, there is increasing evidence that the assumption of rate constancy is often violated, and that DNA and amino acid sequences of even closely related species can evolve at different rates (Bromham and Penny, 2003; Wu and Li, 1985). The reasons given for these deviations from the clock-like model of sequence evolution are related to generation time (Ohta, 2002), metabolic rate (Martin and Palumbi, 1993), mutation rate (Ota and Penny, 2003), and the effect of effective population size on the rate of fixation of mutations (Ohta, 2002). In practice, clock-like behaviour of the data can be tested using a LRT statistic (see above). If the null hypothesis of a constant rate is rejected, the use of methods that try to model rate changes over the tree (so-called "relaxed clock methods") is necessary. There are many such methods that use different approaches to either correct or incorporate rate heterogeneity in the dating process on the basis of specific rate change models (Bromham and Penny, 2003; Felsenstein, 2004). Of all these methods, three are becoming increasingly popular: nonparametric rate smoothing (Sanderson, 1997), penalized likelihood (Sanderson, 2002), and Bayesian rate autocorrelation dating (Kishino et al., 2001; Thorne and Kishino, 2002; Thorne et al., 1998).

Nonparametric rate smoothing attempts to simultaneously estimate unknown divergence times and smooth the rapidity of change along lineages. To smooth rate changes, a nonparametric function is used that penalizes rates that change too fast from branch to neighbouring branch, thus reflecting an idea of autocorrelation of rates. Because the penalty function includes unknown times, an optimality criterion based on this penalty permits an estimation of the divergence times (Sanderson, 1997). Penalized likelihood is a semi-parametric technique that combines likelihood and the nonparametric penalty function used in nonparametric rate smoothing. It permits the specification of the relative contribution of the rate smoothing and the data-fitting parts of the estimation procedure. The optimal level of smoothing can be estimated by running a cross-validation procedure (Sanderson, 2002). Both nonparametric rate smoothing and penalized likelihood methods provide confidence intervals on the estimated parameters based on two alternative strategies (using the curvature of the likelihood surface around the parameter estimate, and calculating an age distribution based on

chronograms generated from bootstrapped datasets), and allow multiple calibration constraints to permit scaling of rates and times to real units (Sanderson, 2003).

The Bayesian rate autocorrelation dating uses a fully probabilistic and high parametric model to describe the change in evolutionary rate over time, and uses MCMC approximation to derive the posterior distribution of rates and times from a prior distribution. For the assignments of rates to different branches in the tree, rates are drawn from a lognormal distribution, and a parameter called Brownian motion constant describes the amount of autocorrelation (Kishino et al., 2001; Thorne and Kishino, 2002; Thorne et al., 1998). In order to scale rates and times, the prospective age of the root node must be specified *a priori*. This method provides Bayesian credibility intervals for estimated divergence times and substitution rates, and allows multiple calibration constraints on nodes (specified as prior age intervals). In contrast to nonparametric rate smoothing and penalized likelihood methods, Bayesian dating method is able to account for multiple genes/loci (or dataset partitions in general) with different evolutionary behaviours. This simultaneous analysis of multiple genes may yield more accurate estimates of divergence times (Thorne and Kishino, 2002).

1.3.5. *Molecular Markers*

Of the various molecular markers that have been employed in phylogenetic studies (see Hillis et al., 1996; Rokas and Holland, 2000 for a review), the analysis of DNA (and/or amino acid) sequences has far become the most widely used nowadays, particularly since the advent of the polymerase chain reaction (PCR; Saiki et al., 1988). It is now possible and relatively easy to determine the precise nucleotide sequence of specific genes (or sets of genes) for entire groups of organisms, and use that information to get deep insight into their phylogenetic relationships and molecular evolution. However, the choice of specific genes that are most appropriate for the phylogenetic question at hand is a crucial step, as the results of the study are largely dependent on it. In general, the use of "favourite" genes or genomic regions in phylogenetic studies is more commonly related to the technical ease with which their sequences can be determined, and their "success" in previous similar-level studies. Over the years, ribosomal genes (particularly mitochondrial ones) have long been used in animal phylogenetic studies at various taxonomic levels. Also, some mitochondrial protein-coding genes, such as *cob* and *cox1*, have become particularly popular. All these genes grew in popularity often because the early availability of "universal" PCR primers for them (Kocher et al., 1989; Palumbi et al., 1991), but in most cases, only short partial fragments (300-600

base pairs [bp]) of these genes are sequenced. Most of the "favourite" genes are encoded by mitochondrial DNA because some of its features (lack of introns, maternal inheritance, practical absence of recombination, and haploidy) have made it particularly adequate for estimating animal molecular phylogenies (reviewed by Avise, 1994; Meyer, 1993).

In recent years, several studies have demonstrated the need to establish high-level phylogenetic inferences based on rather large sequence datasets in order to achieve statistical confidence (Cummings et al., 1995; Russo et al., 1996; Zardoya and Meyer, 1996). There is a growing trend to sequence big genomic regions to tackle phylogenetic problems, and, in particular, sequencing and analysis of complete mitochondrial genomes is providing reliable estimations of deep phylogenies in many animal groups, and, for instance, has allowed inferring evolutionary relationships among the three orders of living amphibians and other groups of tetrapods (Zardoya and Meyer, 2001).

The animal mitochondrial genome is almost always a circular molecule (but see Bridge et al. [1992] and Raimond et al. [1999] for exceptions in cnidarians and a crustacean) of about 16000 bp, typically containing the same set of 37 genes, encoding 13 protein subunits of the enzymes of oxidative phosphorilation, two ribosomal ribonucleic acids (RNAs) of the mitochondrial ribosome, and 22 transfer RNAs (tRNAs) necessary for the translation of the proteins encoded by mitochondrial DNA (Boore, 1999; Jameson et al., 2003; Wolstenholme, 1992) (Fig. 8). There is also generally a single large non-coding region that may contain controlling elements for replication and transcription of the mitochondrial genome (Shadel and Clayton, 1997). In the vertebrate mitochondrial DNA, this large non-coding region (dubbed "control region") contains the origin of heavy-strand replication, and there is also a short non-coding region containing the origin of light-strand replication (Shadel and Clayton, 1997) (Fig. 8). Several methods exist to determine the nucleotide sequence of the complete mitochondrial genome of particular taxa. Among them, the method employed in the studies presented in this thesis consists in the use of a set of primers to amplify by PCR contiguous, overlapping fragments that cover the entire mitochondrial DNA. The resulting PCR products are purified and cloned into vectors, and the recombinant plasmids are eventually sequenced in an automated DNA sequencer. A more detailed description of the methodology employed is given in the "Materials and methods" section of Publication II of this thesis. A few of the employed primer sequences were taken from the literature, but most of them were designed *de novo* in conserved mitochondrial regions using an alignment for amphibian taxa.

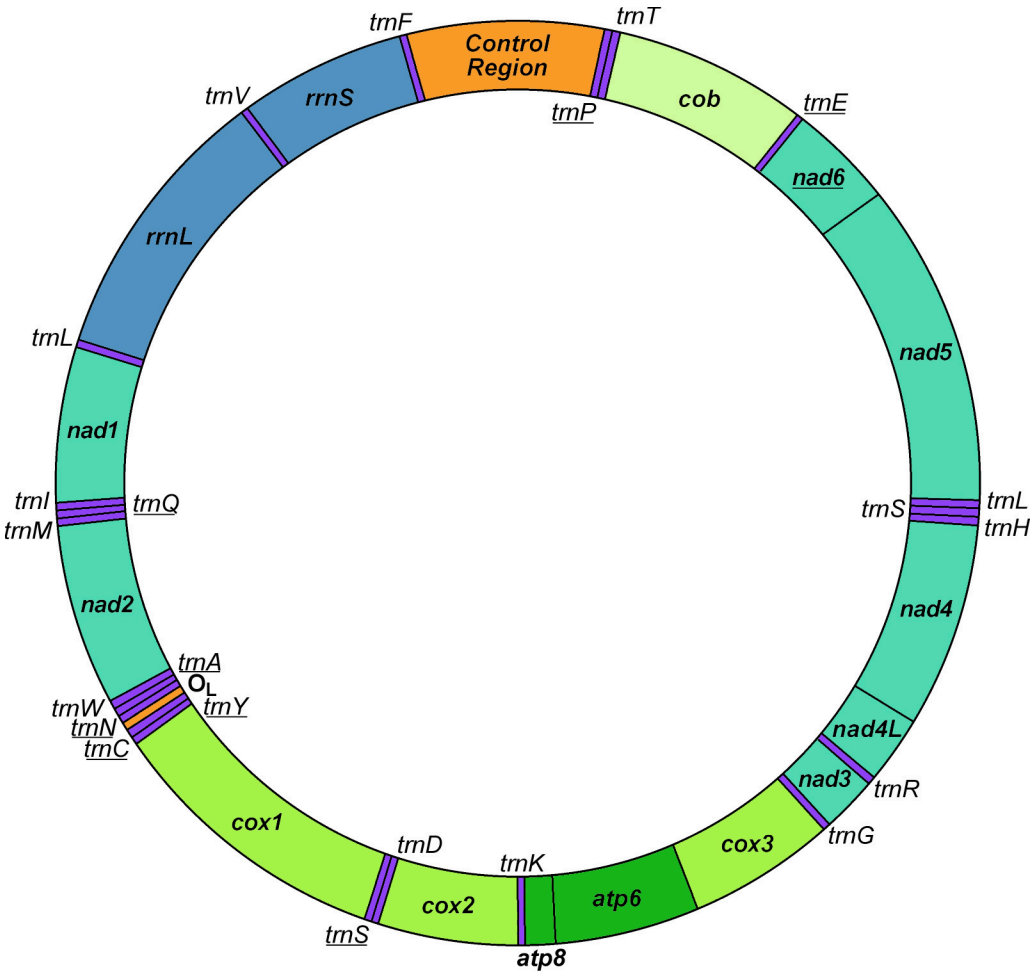


Fig. 8. Consensus organization of the vertebrate mitochondrial genome (Boore, 1999; Jameson et al., 2003). Genes encoded by the light-strand are underlined. Abbreviations: *atp6-8*, genes for ATP synthase F0 subunits 6 and 8; *cox1-3*, genes for cytochrome c oxidase subunits 1-3; *cob*, gene for cytochrome b; *nad1-6*, genes for NADH dehydrogenase subunits 1-6; *rrnS* and *rrnL*, genes for small and large subunits of ribosomal RNA; *trn*, transfer RNA gene (each abbreviated by the corresponding one-letter amino acid code); O_L, origin of light-strand replication.

Several recent studies have demonstrated that some orthologous nuclear protein-coding genes outperform mitochondrial sequences in reconstructing ancient relationships (Groth and Barrowclough, 1999; Springer et al., 2001). One of these single-copy nuclear protein-coding genes that have proven useful in inferring deep relationships among major tetrapod lineages (Groth and Barrowclough, 1999; Hoegg et al., 2004; Murphy et al., 2001; Townsend et al., 2004) is the Recombination Activating Gene 1 (*rag1*). This gene is required for V(D)J recombination (V, D, and J being, respectively, the "Variable", "Diversity", and "Joining" segments of the genes encoding the variable portion of the T-cell antigen receptors), which is part of the adaptative (antigen-specific) immune response of vertebrates (Agrawal et

al., 1998; Schatz et al., 1989), and it appears it was originally part of a transposon that entered the genome at the time jawless and jawed vertebrates separated (Agrawal et al., 1998; Thompson, 1995). A number of properties favouring the phylogenetic utility of *rag1* in vertebrates have been reported: relatively slow substitution rate, lack of intervening introns (at least in tetrapods), rarity of indels, minimal saturation of transition changes at third positions of codons, nearly constant base composition across taxa, and symmetry in directional patterns of reconstructed change (Groth and Barrowclough, 1999; Martin, 1999). The method employed in the studies presented in this thesis to determine *rag1* nucleotide sequences was similar to that described for complete mitochondrial genome sequences (see above). In this case, four newly designed primers were used to amplify two contiguous, overlapping fragments that covered a 1500 bp long portion of the 3'-end part of the *rag1* gene.

2. OBJECTIVES

This Ph.D. thesis aims to study the phylogeny and molecular evolution of caecilian amphibians through the following four objectives:

1. To establish an evolutionary timescale for the origin and diversification of living caecilians.

2. To determine the nucleotide sequence of the complete mitochondrial genome and the *rag1* gene of representative species of the six families of caecilians, and use that information to reconstruct a robust phylogeny of the group.

3. To characterize the structure and organization of the mitochondrial genome of caecilians, as well as the genetic mechanisms that may have originated molecular singularities in particular taxa.

4. To investigate the phylogenetic performance and utility of mitochondrial genes and genomes, and of the nuclear *rag1* gene for caecilian systematics.

3. PUBLICATION I

Title: **Initial diversification of living amphibians predated the breakup of Pangaea**

Authors: Diego San Mauro, Miguel Vences, Marina Alcobendas, Rafael Zardoya, Axel Meyer

Status: Published

Year: 2005

Journal: American Naturalist (SCI Impact Factor: 4.476)

Volume: 165

Pages: 590-599

Resumen I (Spanish translation of the abstract of Publication I)

El origen y divergencia de los tres órdenes de anfibios vivos (Anura, Caudata, Gymnophiona) y sus principales linajes es uno de los asuntos más debatidos en evolución de vertebrados. Aquí, presentamos una filogenia molecular robusta basada en el gen nuclear RAG1 así como resultados de reloj molecular usando una variedad de calibraciones independientes alternativas. Nuestros análisis sugieren que el origen y divergencia temprana de los tres órdenes de anfibios vivos se remonta al Paleozoico o Mesozoico inferior, antes de la ruptura de Pangea, y poco después de la divergencia de los peces de aletas lobuladas. El nuevo escenario biogeográfico resultante, las estimas de edades, y la rápida divergencia inferida entre los tres órdenes de Lissamphibia podría explicar la falta de fósiles que representen ancestros posibles o grupos hermano inmediatos de los tres órdenes, así como la paradójica distribución de algunos taxones de anfibios fósiles. Además, la antigua y rápida radiación de los tres órdenes de Lissamphibia probablemente explica por qué las longitudes de las ramas que conectan sus nodos tempranos son particularmente cortas, y por tanto provocan que la inferencia filogenética de las relaciones implicadas sea especialmente difícil.

Notes and Comments

Initial Diversification of Living Amphibians Predated the Breakup of Pangaea

Diego San Mauro,[1,*] Miguel Vences,[2,†] Marina Alcobendas,[1,‡] Rafael Zardoya,[1,§] and Axel Meyer[3,‖]

1. Departamento de Biodiversidad y Biología Evolutiva, Museo Nacional de Ciencias Naturales, Consejo Superior de Investigaciones Científicas, José Gutiérrez Abascal, 2, E-28006 Madrid, Spain;
2. Institute for Biodiversity and Ecosystem Dynamics, Zoological Museum, University of Amsterdam, Mauritskade 61, 1092 AD Amsterdam, The Netherlands;
3. Lehrstuhl für Zoologie und Evolutionsbiologie, Department of Biology, University of Konstanz, 78457 Konstanz, Germany

Submitted July 27, 2004; Accepted January 25, 2005; Electronically published March 17, 2005

Online enhancements: appendixes.

ABSTRACT: The origin and divergence of the three living orders of amphibians (Anura, Caudata, Gymnophiona) and their main lineages are one of the most hotly debated topics in vertebrate evolution. Here, we present a robust molecular phylogeny based on the nuclear RAG1 gene as well as results from a variety of alternative independent molecular clock calibrations. Our analyses suggest that the origin and early divergence of the three living amphibian orders dates back to the Palaeozoic or early Mesozoic, before the breakup of Pangaea, and soon after the divergence from lobe-finned fishes. The resulting new biogeographic scenario, age estimate, and the inferred rapid divergence of the three lissamphibian orders may account for the lack of fossils that represent plausible ancestors or immediate sister taxa of all three orders and the heretofore paradoxical distribution of some amphibian fossil taxa. Furthermore, the ancient and rapid radiation of the three lissamphibian orders likely explains why branch lengths connecting their early nodes are particularly short, thus rendering phylogenetic inference of implicated relationships especially difficult.

* E-mail: diegos@mncn.csic.es.

† E-mail: vences@science.uva.nl.

‡ E-mail: marina@mncn.csic.es.

§ E-mail: rafaz@mncn.csic.es.

‖ E-mail: axel.meyer@uni-konstanz.de.

Am. Nat. 2005. Vol. 165, pp. 590–599. 0003-0147/2005/16505-40546$15.00.

Keywords: amphibian evolution, Pangaea breakup, molecular phylogeny, molecular clock, multiple calibrations, RAG1.

Living amphibians (Lissamphibia) are a successful and highly diversified group of vertebrates that includes thousands of forms (5,770 species; AmphibiaWeb, January 26, 2005; http://www.amphibiaweb.org/) distributed throughout most habitats in all continents except Antarctica (Duellman and Trueb 1994). They experienced a long evolutionary history dating back at least to the early Triassic, the earliest known fossils being *Triadobatrachus* from Madagascar (Rage and Rocek 1989) and *Czatkobatrachus* from Poland (Evans and Borsuk-Bialynicka 1998). The Lissamphibia are widely thought to be a monophyletic group, constituted by three monophyletic orders (Anura, Caudata, and Gymnophiona) whose origin and interrelationships remain hotly debated (see Meyer and Zardoya 2003 for a recent review). The poor fossil record of some major lissamphibian groups and the fact that the three living amphibian orders possibly acquired their specialized morphology very early in their evolutionary histories (Zardoya and Meyer 2001) have left many questions unresolved regarding the origins, relationships, and historical distribution of the Lissamphibia.

A recent molecular phylogeny of lissamphibians based on mitochondrial rRNA genes grouped salamanders and caecilians to the exclusion of frogs and suggested that the early evolutionary history of living amphibians was associated with the Mesozoic continental fragmentation of the supercontinent Pangaea (Feller and Hedges 1998). Paradoxically, some distributional patterns and some data from the fossil record (Estes and Wake 1972; Estes and Reig 1973; Rage and Rocek 1989; Jenkins and Walsh 1993; Duellman and Trueb 1994; Evans et al. 1996; Evans and Borsuk-Bialynicka 1998; Rocek 2000) point at an initial divergence of living amphibians much earlier than the Mesozoic continental fragmentation of the Pangaea supercontinent. Moreover, alternative molecular phylogenies based on complete mitochondrial genomes (Zardoya and

Meyer 2001; San Mauro et al. 2004) support the "Batrachia" hypothesis (Anura + Caudata).

In order to test whether lissamphibian splits were triggered by Mesozoic continental breakup events, and to distinguish among competing hypotheses, we reconstructed a robust molecular phylogeny based on the RAG1 gene, encompassing for the first time a wide taxon sampling of major lissamphibian lineages. We applied a multiple-calibration Bayesian approach to estimate divergence times. This method was developed to avoid biases that were detected in traditional global molecular clock dating methods (Rodríguez-Trelles et al. 2002; Benton and Ayala 2003). It does not require the assumption of a constant rate of evolution, admits several independent calibrations, and allows the use of prior constraints on divergence time instead of fixed time points (Douzery et al. 2004). To confirm the reliability of the estimates based on the Bayesian relaxed molecular clock dating method, we further provide an empirical comparison of age estimates of basal nodes in the Lissamphibia obtained with a variety of alternative independent molecular clock calibrations (both single and multiple).

Material and Methods

Taxon Sampling and DNA Sequencing

We analyzed 44 amphibian nucleotide sequences of the 3′ end part of the RAG1 gene. This is a nuclear single-copy protein-coding gene that outperforms mitochondrial genes in reconstructing ancient phylogenies (Groth and Barrowclough 1999). The relative rate of evolution of this gene at the nucleotide level is about 2.5 times slower than that of COI (cytochrome c oxidase subunit I) at the amino acid level (San Mauro et al. 2004). For 22 taxa, the sequences were determined for this study using the primers, conditions, and methods reported in San Mauro et al. (2004). Additionally, the following primers were designed to sequence the fragments in some species in which general primers did not amplify: RAG1.R (5′-GGT GYT TYA ACA CAT CTT CCA TYT CRT A-3′), Sal-RAG1.F (5′-CAC YGG GCG CCA GAT YTT CCA RCC-3′), and Sal-RAG1.R1 (5′-AGG TTC TCA GTG TGG CTC CTG GTG A-3′). All nucleotide sequences reported in this article have been deposited in the GenBank database under accession numbers AY583334–AY583355.

Another 22 amphibian RAG1 sequences were obtained from previous studies (Hoegg et al. 2004; San Mauro et al. 2004). The sequences of eight amniotes were used to root the tree; in addition, the coelacanth was used as outgroup for the molecular clock analysis. A complete list of taxa and their higher classification, voucher specimens, collection localities, and GenBank accession numbers can be found in appendix A in the online edition of the *American Naturalist.*

Phylogenetic Reconstruction and Molecular Clock Calibration

Nucleotide sequences were aligned by hand and only one gapped codon was excluded from the analyses (see app. B in the online edition), yielding an alignment of 1,368 positions (only 891 bp were available for *Leiopelma hochstetteri*). RAG1 sequences showed no severe saturation effects, as judged by plots of pairwise differences (absolute, only transitions, and only transversions) versus corrected sequence divergence (measured as maximum likelihood distance, not shown). The RAG1 alignment was subjected to maximum likelihood (ML; Felsenstein 1981), Bayesian inference (BI; Huelsenbeck et al. 2001), minimum evolution (ME; Rzhetsky and Nei 1992), and maximum parsimony (MP; Fitch 1971). Maximum likelihood, ME, and MP analyses were carried out with PAUP* version 4.0b10 (Swofford 1998). Bayesian inference analysis was conducted with MrBayes version 3.0b4 (Huelsenbeck and Ronquist 2001). The best-fitting models of nucleotide substitution were selected using Modeltest version 3.6 (Posada and Crandall 1998), following the Akaike Information Criterion (AIC). Maximum likelihood and ME analyses assumed the parameter-rich GTR (Rodríguez et al. 1990) + Γ + I model for all positions. Bayesian inference analyses were also performed using the GTR + Γ + I substitution model, although in this case parameter estimations were independently assessed for each codon position ("unlink" option). Maximum likelihood, ME, and MP analyses were performed using heuristic searches with TBR branch swapping and 10 random stepwise additions of taxa. Nonparametric bootstrapping was used to test the reliabilities of the ML, ME, and MP trees (100 pseudoreplicates for ML, and 1,000 pseudoreplicates for ME and MP). Bayesian inference analyses were performed simulating four simultaneous chains, for a million generations, sampling every 100 generations. Generations sampled before the chain reached stationarity (100,000) were discarded ("burn-in").

Divergence times were determined using a Bayesian approach that incorporates variation of rates of evolution among genes and among lineages (Thorne et al. 1998; Kishino et al. 2001; Thorne and Kishino 2002). We used the ML topology that was inferred based on the RAG1 data set as the starting phylogeny. Branch lengths of the inferred topology and divergence times were estimated using the programs Estbranches and Multidivtime, respectively (http://statgen.ncsu.edu/thorne/). The Bayesian method also requires the specification of prior distributions for parameters. The prior assumption for the mean

and standard deviation of the time of the ingroup root node (rttm) was set to 42 time units, where 1 time unit in this analysis represents 10 million years. This value was obtained based on the split of coelacanth and tetrapod lineages 420 million years ago (mya; Zhu et al. 2001). The standard deviation of the prior distribution was set to its maximum value (equal to the mean) to avoid violation of the definition of a prior. The divergence among diapsids and synapsids (Kumar and Hedges 1998) was used as the main calibration point. Considering the criticism of Graur and Martin (2004), we calibrated this split at 338–288 mya, as proposed by these authors, and, in addition, included multiple internal calibrations within the Lissamphibia as upper and lower time constraints. Four of these internal calibrations were based on fossil record: minimum age of the frogs-salamander split (node 36; see app. C in the online edition) at 230 mya (fossil record of frog ancestor *Triadobatrachus*; Rage and Rocek 1989); minimum age of the split among hynobiid and cryptobranchid salamanders (node 33) at 161 mya (cryptobranchid fossil record; Gao and Shubin 2003); minimum age of the split of pipid frogs from their sister group (node 24) at 140 mya (records of Mesozoic fossil pipids; Rocek 2000); minimum age of the split between *Caudiverbera* and *Lechriodus* (node 5) at 53 mya (fossil records of *Caudiverbera*; Baez 2000). The other four internal calibrations were based on biogeographical events: minimum age of the split among the caecilians *Gegeneophis* and *Geotrypetes* (node 37) at 130 mya (Gondwana fragmentation, separation of India-Seychelles-Madagascar from Africa; Rabinowitz et al. 1983); minimum age of the separation among South American and African pipid frogs (node 21) at 86 mya (separation of Africa and South America; Pitman et al. 1993); minimum age of the split between *Agalychnis* and *Litoria* (node 1) at 42 mya (last connection between Australia and South America; Seddon et al. 1998); maximum age of the split between *Mantidactylus wittei* and *Mantidactylus* sp. from the Comoro islands (node 7) at 15 mya (volcanic origin of the oldest Comoro island Mayotte; Vences et al. 2003). These calibrations exhibited a significant fit between time and divergence (see app. D in the online edition).

Divergence times were also independently reestimated using the following single and multiple calibrations (see table 1 for details): (1) the single calibration proposed by Kumar and Hedges (1998), (2) the correction to 1 proposed by Graur and Martin (2004), (3) the single calibration proposed by Reisz and Müller (2004), (4) the single calibration used by Vences et al. (2003), and (5) our multiple calibration plus the single calibration proposed by Reisz and Müller (2004).

The Markov chain Monte Carlo (MCMC) method was employed to approximate both prior and posterior distributions (Kishino et al. 2001). Initial parameter values were randomly selected to initialize the Markov chain, and then a burn-in period of 100,000 cycles was completed before parameters were sampled from the MCMC chain. Afterward, the state of the Markov chain was sampled every 100 cycles until a total of 10,000 generations.

Results and Discussion

Early Separation of the Three Lissamphibian Orders during the Paleozoic

According to our results, the ancestral lineage of caecilians separated from the common ancestor of batrachians approximately 367 (417–328; 95% confidence interval [CI]) mya (fig. 1). The divergence of salamanders and frogs occurred shortly thereafter, 357 (405–317) mya (fig. 1). Although the "Batrachia" hypothesis is not strongly supported by our results, it can be considered as the best explanation given the available data on the basis that all phylogenetic methods yielded phylograms with this topology (no method recovered alternative arrangements) and ME and MP found substantial (>70%) statistical support for the clade Batrachia (see also Zardoya and Meyer 2001).

Analyses of our data set with single and alternative calibrations (e.g., those of Kumar and Hedges [1998] and Reisz and Müller [2004]) produced concordant results (table 1). In all cases, a Paleozoic age of separation between the three amphibian orders was estimated (367–297 mya). In addition, all estimates agreed that the initial splittings within living salamanders and frogs occurred during the Permian–Triassic (273–204 mya), whereas the basal splits among living caecilians were estimated to be slightly younger in some of the analyses (214–150 mya).

These results may indicate that the separation of the three orders of modern amphibians in the Paleozoic occurred almost immediately (in evolutionary time) after the "jump to land" of sarcopterygian fishes (360 mya), as had been postulated by Benton (1990), Milner (1993), and Carroll et al. (2004), and in parallel with the diversification of extinct lineages of amphibians (e.g., *Acanthostega* or *Ichthyostega*). Such a rapid radiation event may be the cause for the lack of fossils that represent plausible ancestors or morphological immediate sister taxa of all three lissamphibian orders and the particularly short branch lengths connecting the nodes among them, thereby rendering phylogenetic inference more difficult.

These results disagree with the hypothesis that salamanders (Laurasia) and caecilians (Gondwana) arose in the Mesozoic from a common ancestor by vicariance directly linked to the breakup of supercontinent Pangaea, with frogs separating from the amphibian stem lineage much earlier during the Paleozoic (Feller and Hedges

Table 1: Comparison of age estimates of basal nodes in the Lissamphibia, their standard deviation (SD), and 95% confidence intervals (CI) obtained with different calibrations (multiple and single)

Node and calibration[a]	Node[b]	Age	SD	Upper CI	Lower CI
Gymnophiona-Batrachia:					
Multiple	43	367	23	328	417
Kumar and Hedges 1998	43	342	16	315	376
Graur and Martin 2004	43	344	22	305	392
Reisz and Müller 2004	43	359	39	299	453
Vences et al. 2003	43	309	101	144	534
Multiple + RM	43	366	23	325	416
Caudata-Anura:					
Multiple	36	357	22	317	405
Kumar and Hedges 1998	36	329	17	297	365
Graur and Martin 2004	36	331	23	289	379
Reisz and Müller 2004	36	346	38	285	436
Vences et al. 2003	36	297	98	137	514
Multiple + RM	36	356	22	315	405
Gymnophiona:					
Multiple	42	214	20	177	256
Kumar and Hedges 1998	42	168	27	115	221
Graur and Martin 2004	42	169	28	115	224
Reisz and Müller 2004	42	177	30	121	239
Vences et al. 2003	42	150	58	61	281
Multiple + RM	42	213	20	177	254
Caudata:					
Multiple	35	273	19	238	312
Kumar and Hedges 1998	35	229	23	182	273
Graur and Martin 2004	35	231	26	180	280
Reisz and Müller 2004	35	241	32	184	313
Vences et al. 2003	35	206	72	90	365
Multiple + RM	35	271	19	237	312
Anura:					
Multiple	24	262	21	223	305
Kumar and Hedges 1998	24	227	22	184	268
Graur and Martin 2004	24	228	24	180	276
Reisz and Müller 2004	24	238	31	183	307
Vences et al. 2003	24	204	70	91	359
Multiple + RM	24	262	21	222	305
Hyloidea:					
Multiple	4	65	8	52	84
Kumar and Hedges 1998	4	42	10	26	63
Graur and Martin 2004	4	42	10	25	64
Reisz and Müller 2004	4	44	11	26	68
Vences et al. 2003	4	37	15	15	72
Multiple + RM	4	65	8	52	84
Ranoidea:					
Multiple	9	99	16	70	132
Kumar and Hedges 1998	9	78	16	50	111
Graur and Martin 2004	9	78	16	50	113
Reisz and Müller 2004	9	82	18	52	121
Vences et al. 2003	9	69	25	28	127
Multiple + RM	9	99	16	71	132

[a] The nodes refer to the splits between caecilians and the salamander-frog clade (Batrachia), between salamanders and frogs, and to the initial splits of caecilians, salamanders, frogs, hyloids, and ranoids. The calibrations used were (1) the preferred multiple calibration as described in "Material and Methods" and shown in figure 1; (2) the single calibration proposed by Kumar and Hedges (1998), namely, a fixed synapsid-diapsid divergence at 310 mya; (3) the correction to the synapsid-diapsid calibration proposed by Graur and Martin (2004), 288–338 mya; (4) the single calibration proposed by Reisz and Müller (2004) for the crocodile-bird split, 227–242 mya; (5) the calibration used by Vences et al. (2003) based on endemic frogs of the oceanic island Mayotte (maximum age constraint 15 mya); (6) the preferred multiple calibration plus the single calibration proposed by Reisz and Müller (2004; "Multiple + RM").

[b] Node numbers are as in appendix C in the online edition.

1998). That hypothesis was based on a ribosomal molecular phylogeny and the geographical distribution of the amphibian fossil record, but it lacked molecular clock estimates. The RAG1-based hypothesis of a Paleozoic origin of all modern amphibian groups predating the breakup of Pangaea, as well as the tentative salamander + frog clade in our tree, therefore invalidate Feller and Hedges's (1998) hypothesis. Furthermore, the presence of the putative stem-group caecilian *Eocaecilia* in Laurasia (early Jurassic of North America; Jenkins and Walsh 1993) could not previously be reconciled with that hypothesis.

Initial Splittings within the Living Caecilians in the Early Mesozic

The presence of living caeciliids in South America, Africa, Seychelles, and India, as well as the African affinities of a Paleocene caeciliid fossil (*Apodops*) found in South America (Estes and Wake 1972) suggest that the split of the major extant caecilian lineages occurred before the breakup of Gondwana. A successive dispersal from the Indian Plate subsequent to its collision with Asia has been proposed to explain the origin of ichthyophiid caecilians in Southeast Asia (Duellman and Trueb 1994; Wilkinson et al. 2002). Our results indicate that the time of initial splitting within the modern caecilians occurred about 214 (256–177) mya (fig. 1), when the rhinatrematid lineage separated from the ancestry of all other caecilians, and that the main basal divergences (including the time of initial splitting within the higher caecilians comprising scolecomorphids, caeciliids, and typhlonectids 177 [218–148] mya) took place in the early Mesozoic (fig. 1). Both the old origin, before the breakup of Gondwana, and the presently restricted geographical distribution of many caecilian lineages may indicate that the most ancient clades (rhinatrematids and the ichthyophiid + uraeotyphlid clade) are relicts of groups that may once have been widespread in Gondwana, whereas more recently derived clades such as scolecomorphids and typhlonectids may have evolved in situ and never achieved a wider distribution (Duellman and Trueb 1994).

Initial Splittings within the Living Salamanders in the Late Paleozoic

Salamanders have a mostly Laurasian distribution, and it seems fairly clear that all salamander lineages arose in the Laurasian part of Pangaea (Duellman and Trueb 1994). However, Mesozoic sirenid fossils are known from both South America (*Noterpeton*) and Africa (*Kababisha*) (Evans et al. 1996) and may raise doubts about an exclusive Laurasian origin of salamanders. Our results indicate that the initial splitting within modern salamanders occurred during the late Paleozoic, 273 (312–238) mya, when the sirenids and the hynobiid + cryptobranchid clade separated from the ancestor of all other salamanders (fig. 1). Interestingly, cryptobranchids, hynobiids, and sirenids all have external fertilization and angular and prearticular bones of the lower jaw not fused, which are considered ancestral traits (Duellman and Trueb 1994). The estimated time of separation of the plethodontids from the ambystomatid + salamandrid clade later occurred about 253 (294–213) mya, and of the ambystomatids from salamandrids about 230 (274–188) mya (fig. 1). Hence, the main divergences of salamanders must have taken place before the breakup of Pangaea and also before the earliest fragmentation of Laurasian landmasses, which began with the opening of the North Atlantic Ocean in the early Jurassic (Smith et al. 1994).

Initial Splittings within the Living Frogs in the Late Paleozoic

The discoveries of *Triadobatrachus* from the early Triassic of Madagascar (Rage and Rocek 1989) and *Czatkobatrachus* from the early Triassic of Poland (Evans and Borsuk-Bialynicka 1998) suggest that Salientia (the stem group of frogs) occurred in all Pangaea. Duellman and Trueb (1994) considered the leiopelmatids to be the sister group of all other frogs, widely distributed before the breakup of Pangaea (Jurassic fossils, *Vieraella* and *Notobatrachus*, are known from Argentina; Estes and Reig 1973), of which the living genera (*Ascaphus* in North America and *Leiopelma* in New Zealand) are merely relicts. Our results show that the estimated time of initial splitting within the living frogs occurred about 263 (305–223) mya, when the leiopelmatids separated from the ancestor of all other living frogs (fig. 1). The subsequent estimated dates of origin of pipids at about 245 (288–204) mya, discoglossids at 235 (277–195) mya, and pelobatoideans 216 (260–176) mya indicate that the divergences of all major archaeobatrachian groups occurred much earlier than the Pangaean fragmentation (fig. 1). These age estimates, together with the recovered paraphyly of archaeobatrachians, may indicate that they are likely remnants of an ancient and relatively fast radiation (Duellman and Trueb 1994; Hoegg et al. 2004) and would call into question the earlier proposal (Feller and Hedges 1998) of a Mesozoic vicariant origin of archaeobatrachians and neobatrachians being directly linked to the fragmentation of Pangaea. Furthermore, the present and Mesozoic fossil Gondwanan distribution of pipid frogs (Duellman and Trueb 1994; Rocek 2000) is geographically inconsistent with that proposal (Feller and Hedges 1998).

Most of the neobatrachian families sampled in this study were clearly placed in either of two well-defined clades,

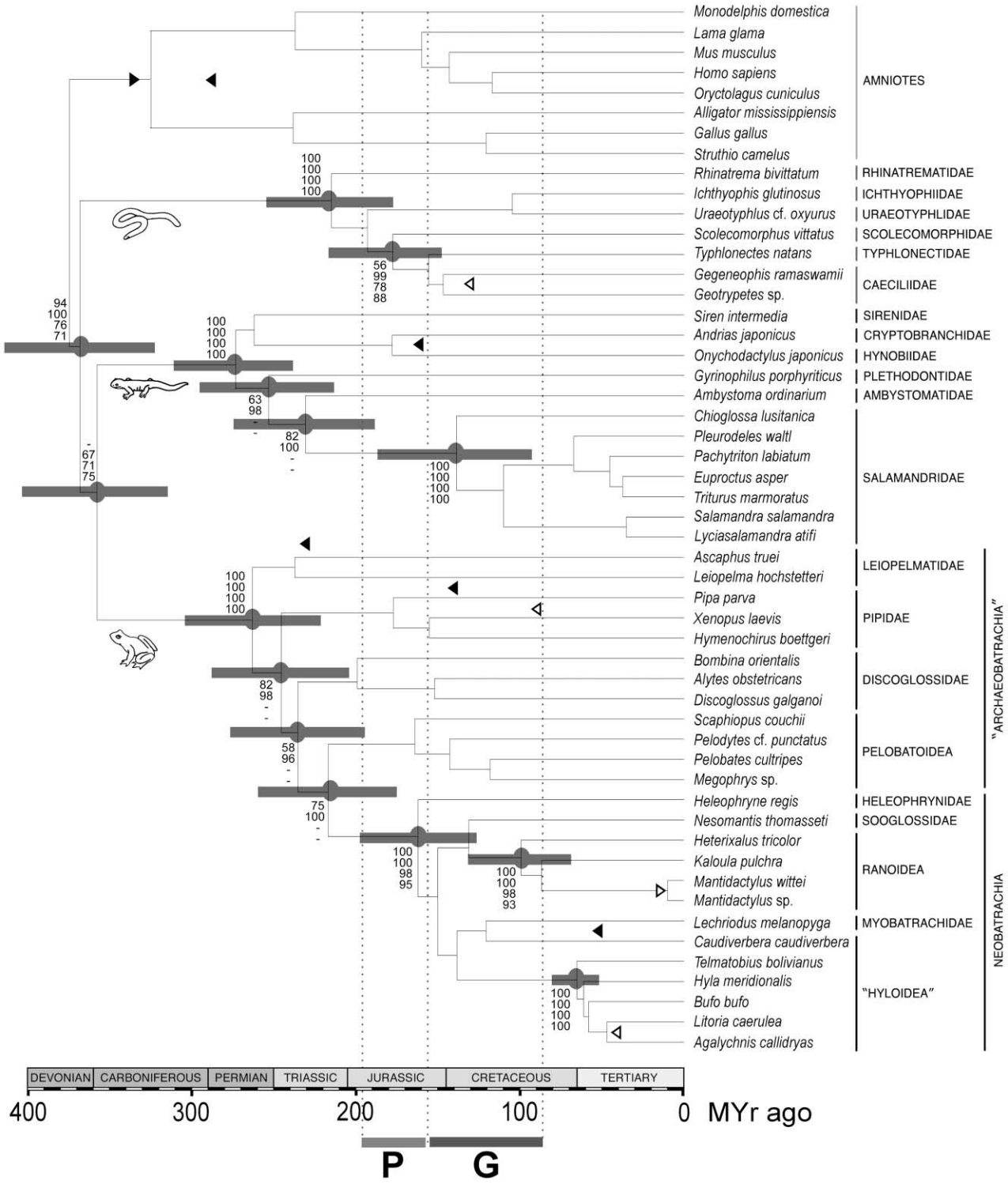


Figure 1: Maximum likelihood phylogeny and estimates of time divergence for the major lineages of living amphibians, estimated from 1,368 nucleotide positions of the RAG1 gene. Calibrations, as listed in "Material and Methods," are marked by triangles (upper and lower bounds). Filled triangles represent calibrations based on fossil record; open triangles represent calibrations based on biogeography. Gray bars mark confidence intervals of age estimates. Dotted vertical lines mark the periods of the initial breakup of Pangaea in Laurasia and Gondwana (*P*) and the breakup of Gondwana (*G*). Numbers adjacent to nodes indicate support for maximum likelihood (upper value of each quartet), Bayesian inference (middle-upper value), minimum evolution (middle-lower value), and maximum parsimony (lower value). Hyphens indicate support values of <50. Statistical support and confidence intervals are shown only for nodes relevant to the "Discussion." A detailed table with support values and age estimates for all nodes can be found in appendix C in the online edition.

the Hyloidea and Ranoidea, which are species-rich radiations containing many additional families (Hoegg et al. 2004). These radiations, according to our molecular clock estimates, occurred at around 65 (83–52) mya and 99 (132–70) mya, respectively (fig. 1), which agrees well with the hypothesis (Feller and Hedges 1998) that they occurred in South America and Africa after the separation of these two continents at 110–86 mya (Pitman et al. 1993). Their young age is not an artifact caused by any of the internal calibrations because calculations based on the single synapsid-diapsid split (table 1) led to similar estimates. Interestingly, the leptodactylid *Caudiverbera* is strongly recovered outside the clade comprising all other Hyloidea (fig. 1; see also app. B), suggesting that its family attribution needs to be revised.

It is remarkable that several species-poor neobatrachian clades originated in much earlier periods than the radiations of hyloideans and ranoideans. This includes the South African heleophrynids, the Australian myobatrachids, the Seychellean sooglossids, and, according to our data, the Neotropical *Caudiverbera*, which is restricted to the southern tip of South America. Probably also the recently discovered Indian *Nasikabatrachus* is one of these early lineages of neobatrachians (Biju and Bossuyt 2003) that radiated, according to our new data, between about 162 (199–128 mya; split of *Heleophryne* from other neobatrachians) and 120 (154–91 mya; split of *Caudiverbera* from myobatrachids) mya. This initial diversification occurred before the breakup of Gondwana, as indicated by the wide, though localized, distribution of their extant representatives. Their current restriction to geographic refuges indicates that these early neobatrachians may have been more widespread but were outcompeted by the more modern hyloid and ranoid radiations in large parts of their original distribution area.

Reliability of Relaxed Clock Estimates

Until recently, molecular datings were estimated under the assumption of a constant-rate evolution (Nei et al. 2001). To estimate divergence times, a linearized (ultrametric) tree was constructed, and a timescale for the tree was produced using one or several (through a linear regression fitting) calibration points. Molecular clocks estimated this way are highly controversial because they often conflict with paleontological evidence (Benton and Ayala 2003). The source of this discrepancy relies on constraints inherent to both kinds of data. Divergence times inferred by paleontologists can only be underestimates of the actual origin of a lineage (Benton and Ayala 2003), provided that chronological assignments of fossils are correct. Moreover, if the fossil record for a given lineage is particularly poor, these underestimations can become misleading (Reisz and Müller 2004). On the other hand, conventional molecular dating approaches suffer from several limitations that lead to overestimation biases (Rodríguez-Trelles et al. 2002; Benton and Ayala 2003). Limited taxon sampling or calibration points can seriously affect molecular dating estimates (Douzery et al. 2004). However, the most pervasive handicaps are the significant violations of the assumption of a constant rate of evolution that may be undetected due to the limited statistical power of relative-rate tests (Bromham et al. 2000) and the asymmetric distribution of molecular time estimates (with an unconstrained older end) that leads to a systematic overestimation bias (Rodríguez-Trelles et al. 2002). Well-known examples of this controversy (i.e., consistently older molecular estimates than known fossil evidence) have been reported at the origin of vascular land plants, modern birds, and placental mammals (Benton and Ayala 2003).

In this study, we have tried to reduce the biases of conventional molecular dating by selecting a gene that has an appropriate rate of evolution for the question at hand, increasing taxon sampling, and applying the most recent Bayesian analytical techniques that relax molecular clock assumptions and allow the incorporation of multiple independent calibration constraints. A recent study (Douzery et al. 2004) showed that estimated molecular ages using the same Bayesian approach are less prone to overestimation than conventional molecular clock methods. Therefore, we believe that most of our molecular age estimates can be considered a reasonable approximation of the actual divergence times for the main lineages of living amphibians. Indeed, many molecular date estimates within the lissamphibian clade seem to agree very well with paleontological evidence. For instance, recent paleontological studies place the separation of the three orders of living amphibians back into the early Carboniferous (Carroll 2001; Carroll et al. 2004). The means of our estimated dates for these splits go back into the late Devonian, but CIs of these estimates also cover the early Carboniferous. Therefore, we cannot rule out a slight overestimation that is negligible when CIs are considered. Nonetheless, we are aware that some dates may be considerably overestimated, as is the case for example of the split between marsupials and placental mammals. This divergence is thought to have occurred sometime in the late Jurassic (Kumar and Hedges 1998) or early Cretaceous (Benton 1990). However, our analyses place this divergence between the late Permian and early Jurassic. Although the source of this discrepancy is unclear, it may be related to the limited taxon sampling within the outgroup.

Although we believe that most of our time estimates are most likely quite accurate, we are aware that they need to be interpreted with caution. In any case, overall the estimated dates for the initial splits within the living am-

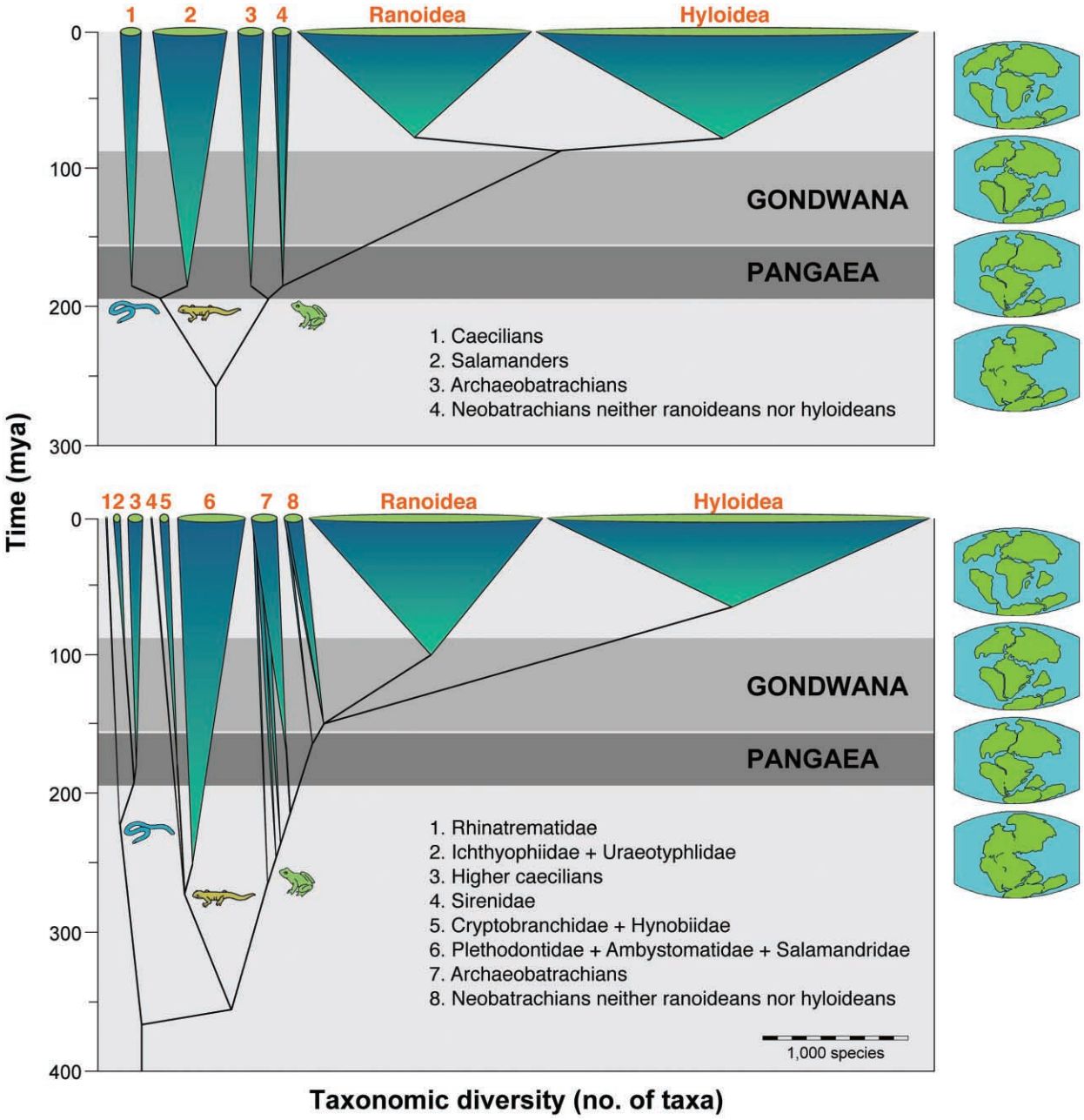


Figure 2: Comparative schematic graph of the radiations of living amphibians according to results in this study (*bottom*) and Feller and Hedges's (1998) hypothesis (*top*). The cross sections of the cones indicate roughly the number of extant species within a group. Shaded bands mark the periods of the breakups of Pangaea and Gondwana.

phibians are so old in comparison with the breakup of Pangaea that it is rather improbable that these splits were actually linked to the continental fragmentation of this supercontinent. Additional information from other genes (Nei et al. 2001; Thorne and Kishino 2002) and fossils (Reisz and Müller 2004) as well as finer calibrations would be desirable to obtain more accurate time estimates and would help reconcile molecular and fossil evidence.

Conclusions

This study presents a comprehensive sampling of most major amphibian lineages for a nuclear protein-coding

gene, and it is the first that makes use of multiple and independent calibrations across the different lissamphibian groups to date major cladogenetic events within extant amphibians. Our results reject the hypothesis that early lissamphibian diversification was triggered by the continental breakup of Pangaea. A few phylogenetic patterns and datings recovered herein agree with scenarios of vicariance in the context of continental breakup, such as the hyloid-ranoid split and the initial diversification of neobatrachians (fig. 2). However, the origin as well as the initial diversification of salamanders, frogs, and caecilians predated the fragmentation of Pangaea (fig. 2). Antiquity of lissamphibian branches likely accounts for the long independent evolution of many convergent patterns in morphology and life history (Duellman and Trueb 1994). Our data provide old age estimates for many extant lissamphibian groups, but they also suggest that the most diverse clades (hyloid and ranoid neobatrachians, which together contain more species than all other amphibians combined; fig. 2) are younger than commonly thought. Ecological displacement by such young species-rich radiations might therefore have caused the extinction and current geographical restrictions of most older taxa, thereby masking the initial biogeographic patterns. Our study thereby provides a useful evolutionary framework that will be important in future studies on amphibian biology. The hypothesis presented here of a probable old origin of many of the major lineages of living amphibians, some of which are geographically restricted and now species poor, turns them into real "living fossils" among extant tetrapods, emphasizing the importance and urgency of the efforts that should be afforded for their conservation.

Acknowledgments

We are grateful to B. Arano, I. de la Riva, I. Doadrio, K. M. Enge, M. García-París, N. J. Gemmel, K. P. Gomas, S. I. Hoegg, W. B. Love, Í. Martínez-Solano, E. Recuero, S. J. Richards, and T. Ziegler for tissue samples or sequences. P. E. Ahlberg, R. L. Carroll, D. J. Gower, O. Madsen, D. B. Wake, and M. Wilkinson gave insightful comments on this work. M. J. Benton and an anonymous reviewer provided helpful suggestions on an earlier version of the manuscript. D.S.M. was sponsored by a predoctoral fellowship of the Ministerio de Educación y Ciencia of Spain. M.A. was sponsored by a postdoctoral fellowship of the Comunidad de Madrid (Spain). This work received financial support from a project of the Ministerio de Educación y Ciencia of Spain to R.Z. (CGL2004-00401) and of the Deutsche Forschungsgemeinschaft to A.M. and M.V.

Literature Cited

Baez, A. M. 2000. Tertiary anurans from South America. Pages 1388–1401 *in* H. Heatwole and R. L. Carroll, eds. Amphibian biology. Surrey Beatty, Chipping Norton, Australia.

Benton, M. J. 1990. Phylogeny of the major tetrapod groups: morphological data and divergence dates. Journal of Molecular Evolution 30:409–424.

Benton, M. J., and F. J. Ayala. 2003. Dating the tree of life. Science 300:1698–1700.

Biju, S. D., and F. Bossuyt. 2003. New frog family from India reveals an ancient biogeographical link with the Seychelles. Nature 425: 711–714.

Bromham, L., D. Penny, A. Rambaut, and M. D. Hendy. 2000. The power of relative rates tests depends on the data. Journal of Molecular Evolution 50:296–301.

Carroll, R. L. 2001. The origin and early radiation of terrestrial vertebrates. Journal of Paleontology 75:1202–1213.

Carroll, R. L., C. Boisvert, J. Bolt, D. M. Green, N. Philip, C. Rolian, R. Schoch, et al. 2004. Changing patterns of ontogeny from osteolepiform fish through Permian tetrapods as a guide to the early evolution of land vertebrates. Pages 321–343 *in* G. Arratia, M. V. H. Wilson, and R. Cloutier, eds. Recent advances in the origin and early radiation of vertebrates. Pfeil, Munich.

Douzery, E. J. P., E. A. Snell, E. Baptese, F. Delsuc, and H. Philippe. 2004. The timing of eukaryotic evolution: does a relaxed molecular clock reconcile proteins and fossils? Proceedings of the National Academy of Sciences of the USA 101:15386–15391.

Duellman, W. E., and L. Trueb. 1994. Biology of amphibians. Johns Hopkins University Press, Baltimore.

Estes, R., and O. Reig. 1973. The early fossil record of frogs: a review of the evidence. Pages 11–63 *in* J. Vial, ed. Evolutionary biology of the anurans. University of Missouri Press, Columbia.

Estes, R., and M. H. Wake. 1972. The first fossil record of caecilian amphibians. Nature 239:228–231.

Evans, S. E., and M. Borsuk-Bialynicka. 1998. A stem-group frog from the Early Triassic of Poland. Acta Palaeontologica Polonica 43:573–580.

Evans, S. E., A. R. Milner, and C. Werner. 1996. Sirenid salamanders and a gymnophionan amphibian from the Cretaceous of the Sudan. Palaeontology 39:77–95.

Feller, A. E., and S. B. Hedges. 1998. Molecular evidence for the early history of living amphibians. Molecular Phylogenetics and Evolution 9:509–516.

Felsenstein, J. 1981. Evolutionary trees from DNA sequences: a maximum likelihood approach. Journal of Molecular Evolution 17: 368–376.

Fitch, W. M. 1971. Toward defining the course of evolution: minimal change for a specific tree topology. Systematic Zoology 20:406–416.

Gao, K. Q., and N. H. Shubin. 2003. Earliest known crown-group salamanders. Nature 422:424–428.

Graur, D., and W. Martin. 2004. Reading the entrails of chickens: molecular timescales of evolution and the illusion of precision. Trends in Genetics 20:80–86.

Groth, J. G., and G. F. Barrowclough. 1999. Basal divergences in birds and the phylogenetic utility of the nuclear RAG-1 gene. Molecular Phylogenetics and Evolution 12:115–123.

Hoegg, S., M. Vences, H. Brinkmann, and A. Meyer. 2004. Phylogeny and comparative substitution rates of frogs inferred from se-

quences of three nuclear genes. Molecular Biology and Evolution 21:1188–1200.

Huelsenbeck, J. P., and F. R. Ronquist. 2001. MrBayes: Bayesian inference of phylogeny. Bioinformatics 17:754–755.

Huelsenbeck, J. P., F. R. Ronquist, R. Nielsen, and J. P. Bollback. 2001. Bayesian inference of phylogeny and its impact on evolutionary biology. Science 294:2310–2314.

Jenkins, F. A., and D. M. Walsh. 1993. An Early Jurassic caecilian with limbs. Nature 365:246–249.

Kishino, H., J. L. Thorne, and W. J. Bruno. 2001. Performance of a divergence time estimation method under a probabilistic model of rate evolution. Molecular Biology and Evolution 18:352–361.

Kumar, S., and S. B. Hedges. 1998. A molecular timescale for vertebrate evolution. Nature 392:917–920.

Meyer, A., and R. Zardoya. 2003. Recent advances in the (molecular) phylogeny of vertebrates. Annual Review of Ecology, Evolution, and Systematics 34:311–338.

Milner, A. R. 1993. The Paleozoic relatives of lissamphibians. Herpetological Monographs 7:8–27.

Nei, M., P. Xu, and G. Glazko. 2001. Estimation of divergence times from multiprotein sequences for a few mammalian species and several distantly related organisms. Proceedings of the National Academy of Sciences of the USA 98:2497–2502.

Pitman, W. C., III, S. Cande, J. LaBrecque, and J. Pindell. 1993. Fragmentation of Gondwana: the separation of Africa from South America. Pages 15–34 *in* P. Goldblatt, ed. Biological relationships between Africa and South America. Yale University Press, New Haven, CT.

Posada, D., and K. A. Crandall. 1998. MODELTEST: testing the model of DNA substitution. Bioinformatics 14:817–818.

Rabinowitz, P. D., M. F. Coffin, and D. Falvey. 1983. The separation of Madagascar and Africa. Science 220:67–69.

Rage, J., and Z. Rocek. 1989. Redescription of *Triadobatrachus massinoti* (Piveteau, 1936) an anuran amphibian from the early Triassic. Paleontographica Abteilung A Palaeozoologie-Stratigraphie 206:1–16.

Reisz, R. R., and J. Müller. 2004. Molecular timescales and the fossil record: a paleontological perspective. Trends in Genetics 20:237–241.

Rocek, Z. 2000. Mesozoic anurans. Pages 1295–1331 *in* H. Heatwole and R. L. Carroll, eds. Amphibian biology. Surrey Beatty, Chipping Norton, Australia.

Rodríguez, F., J. F. Oliver, A. Marín, and J. R. Medina. 1990. The general stochastic model of nucleotide substitution. Journal of Theoretical Biology 142:485–501.

Rodríguez-Trelles, F., R. Tarrío, and F. J. Ayala. 2002. A methodological bias toward overstimation of molecular evolutionary time scales. Proceedings of the National Academy of Sciences of the USA 99:8112–8115.

Rzhetsky, A., and M. Nei. 1992. A simple method for estimating and testing minimum-evolution trees. Molecular Biology and Evolution 9:945–967.

San Mauro, D., D. J. Gower, O. V. Oommen, M. Wilkinson, and R. Zardoya. 2004. Phylogeny of caecilian amphibians (Gymnophiona) based on complete mitochondrial genomes and nuclear RAG1. Molecular Phylogenetics and Evolution 33:413–427.

Seddon, J. M., P. R. Baverstock, and A. Georges. 1998. The rate of mitochondrial 12S rRNA gene evolution is similar in freshwater turtles and marsupials. Journal of Molecular Evolution 46:460–464.

Smith, A. G., D. G. Smith, and B. M. Funnell. 1994. Atlas of Mesozoic and Cenozoic coastlines. Cambridge University Press, Cambridge.

Swofford, D. L. 1998. PAUP*: phylogenetic analysis using parsimony (*and other methods). Version 4.0. Sinauer, Sunderland, MA.

Thorne, J. L., and H. Kishino. 2002. Divergence time and evolutionary rate estimation with multilocus data. Systematic Biology 51:689–702.

Thorne, J. L., H. Kishino, and I. S. Painter. 1998. Estimating the rate of evolution of the rate of molecular evolution. Molecular Biology and Evolution 15:1647–1657.

Vences, M., D. R. Vieites, F. Glaw, H. Brinkmann, J. Kosuch, M. Veith, and A. Meyer. 2003. Multiple overseas dispersal in amphibians. Proceedings of the Royal Society of London B 270:2435–2442.

Wilkinson, M., J. A. Sheps, O. V. Oommen, and B. L. Cohen. 2002. Phylogenetic relationships of Indian caecilians (Amphibia: Gymnophiona) inferred from mitochondrial rRNA gene sequences. Molecular Phylogenetics and Evolution 23:401–407.

Zardoya, R., and A. Meyer. 2001. On the origin of and phylogenetic relationships among living amphibians. Proceedings of the National Academy of Sciences of the USA 98:7380–7383.

Zhu, M., X. Yu, and P. E. Ahlberg. 2001. A primitive sarcopterygian fish with an eyestalk. Nature 410:81–84.

Editor: Jonathan B. Losos
Associate Editor: Allan Larson

Appendix A from D. San Mauro et al., "Initial Diversification of Living Amphibians Predated the Breakup of Pangaea"

(Am. Nat., vol. 165, no. 5, p. 590)

Voucher Specimens, Localities, and Classification of Taxa Studied

Newly determined amphibian sequences (MNCN/ADN, Museo Nacional de Ciencias Naturales, Spain; SIH, University of Konstanz, Germany). Anura: *Leiopelma hochstetteri* (Leiopelmatidae; New Zealand), *Hymenochirus boettgeri* (Pipidae; pet trade), *Alytes obstetricans* (Discoglossidae; MNCN/ADN 4313; Tielmes, Spain), *Bombina orientalis* (Discoglossidae; MNCN/ADN 4314; pet trade), *Discoglossus galganoi* (Discoglossidae; MNCN/ADN 4315; Reliegos, Spain), *Pelodytes* cf. *punctatus* (Pelodytidae; MNCN/ADN 8000; Portalegre, Portugal), *Lechriodus melanopyga* (Myobatrachidae; MNCN/ADN 8001; pet trade, Papua New Guinea), *Caudiverbera caudiverbera* (Leptodactylidae; MNCN/ADN 8002; pet trade, Chile), *Bufo bufo* (Bufonidae; MNCN/ADN 8003; Valdemanco, Spain), *Hyla meridionalis* (Hylidae; MNCN/ADN 8004; Logrosán, Spain), *Telmatobius bolivianus* (Leptodactylidae; MNCN/ADN 563; La Paz, Bolivia). Caudata: *Andrias japonicus* (Cryptobranchidae; MNCN/ADN 8005; pet trade), *Onychodactylus japonicus* (Hynobiidae; SIH-13), *Siren intermedia* (Sirenidae; Tallahassee, FL, USA), *Gyrinophilus porphyriticus* (Plethodontidae; MNCN/ADN 8006; North Carolina, USA), *Ambystoma ordinarium* (Ambystomatidae; MNCN/ADN 8007; Michoacán, Mexico), *Chioglossa lusitanica* (Salamandridae; MNCN/ADN 8008; Pobra do Caramiñal, Spain), *Salamandra salamandra* (Salamandridae; MNCN/ADN 8009; Miraflores de la Sierra, Spain), *Triturus marmoratus* (Salamandridae; MNCN/ADN 8010; Arrillor, Spain), *Euproctus asper* (Salamandridae; MNCN/ADN 8011; Zuriza, Spain), *Pachytriton labiatum* (Salamandridae; MNCN/ADN 8012; pet trade, China). Gymnophiona: *Geotrypetes* sp. (Caeciliidae; pet trade, Cameroon).

Amphibian sequences from previous studies (with GenBank accession numbers). Anura: *Ascaphus truei* (Leiopelmatidae; AY323754), *Pipa parva* (Pipidae; AY323761), *Xenopus laevis* (Pipidae; L19324), *Pelobates cultripes* (Pelobatidae; AY323758); *Scaphiopus couchii* (Pelobatidae; AY323759), *Megophrys* sp. (Megophryidae; AY323760), *Heleophryne regis* (Heleophrynidae; AY323764), *Nesomantis thomasseti* (Sooglossidae; AY323778), *Heterixalus tricolor* (Hyperoliidae; AY323768), *Mantidactylus* sp. (Mantellidae; AY323775), *Mantidactylus wittei* (Mantellidae; AY323774), *Kaloula pulchra* (Microhylidae; AY323772), *Litoria caerulea* (Hylidae; AY323767), *Agalychnis callidryas* (Hylidae; AY323765). Caudata: *Lyciasalamandra atifi* (Salamandridae; AY456261), *Pleurodeles waltl* (Salamandridae; AJ010258). Gymnophiona: *Rhinatrema bivittatum* (Rhinatrematidae; AY456257), *Ichthyophis glutinosus* (Ichthyophiidae; AY456256), *Uraeotyphlus* cf. *oxyurus* (Uraeotyphlidae; AY456259), *Scolecomorphus vittatus* (Scolecomorphidae; AY456258) *Gegeneophis ramaswamii* (Caeciliidae; AY456255), *Typhlonectes natans* (Typhlonectidae; AY456260).

Outgroup sequences. Alligator mississippiensis (Crocodylidae; AF143724), *Struthio camelus* (Struthionidae; AF143727), *Gallus gallus* (Phasianidae; M58530), *Monodelphis domestica* (Didelphidae; U51897), *Lama glama* (Camelidae; AF305953), *Mus musculus* (Muridae; M29475), *Oryctolagus cuniculus* (Leporidae; M77666), *Homo sapiens* (Hominidae; NM_000448), *Latimeria menadoensis* (Coelacanthidae; AY442925).

Appendix B from D. San Mauro et al., "Initial Diversification of Living Amphibians Predated the Breakup of Pangaea"

(Am. Nat., vol. 165, no. 5, p. 590)

Gapped Codon Excluded from the Alignment

A synapomorphic codon insertion was observed in the RAG1 nucleotide sequence of the representatives of our well-defined Hyloidea clade (*Telmatobius bolivianus*, *Litoria caerulea*, *Agalychnis callidryas*, *Hyla meridionalis*, and *Bufo bufo*) with respect to all other amphibians and amniotes. This indel corresponds to that reported by Venkatesh et al. (2001) between positions 637 and 638 of the human RAG1 amino acid sequence and is an amino acid deletion in tetrapods with respect to fishes. Our more comprehensive alignment allowed us to correctly relocate the indel at amino acid position 636–637, where lobe-finned fishes had a serine that is lost in tetrapods but secondarily reevolved in the above-mentioned anuran species, thereby providing further evidence for the monophyly of the group to the exclusion of *Caudiverbera* (which lacks this synapomorphic trait).

Literature Cited in Appendix B

Venkatesh, B., M. V. Erdmann, and S. Brenner. 2001. Molecular synapomorphies resolve evolutionary relationships of extant jawed vertebrates. Proceedings of the National Academy of Sciences of the USA 98: 11382–11387.

Appendix C from D. San Mauro et al., "Initial Diversification of Living Amphibians Predated the Breakup of Pangaea"

(Am. Nat., vol. 165, no. 5, p. 590)

Extended Result of the Phylogenetic and Molecular Clock Analyses

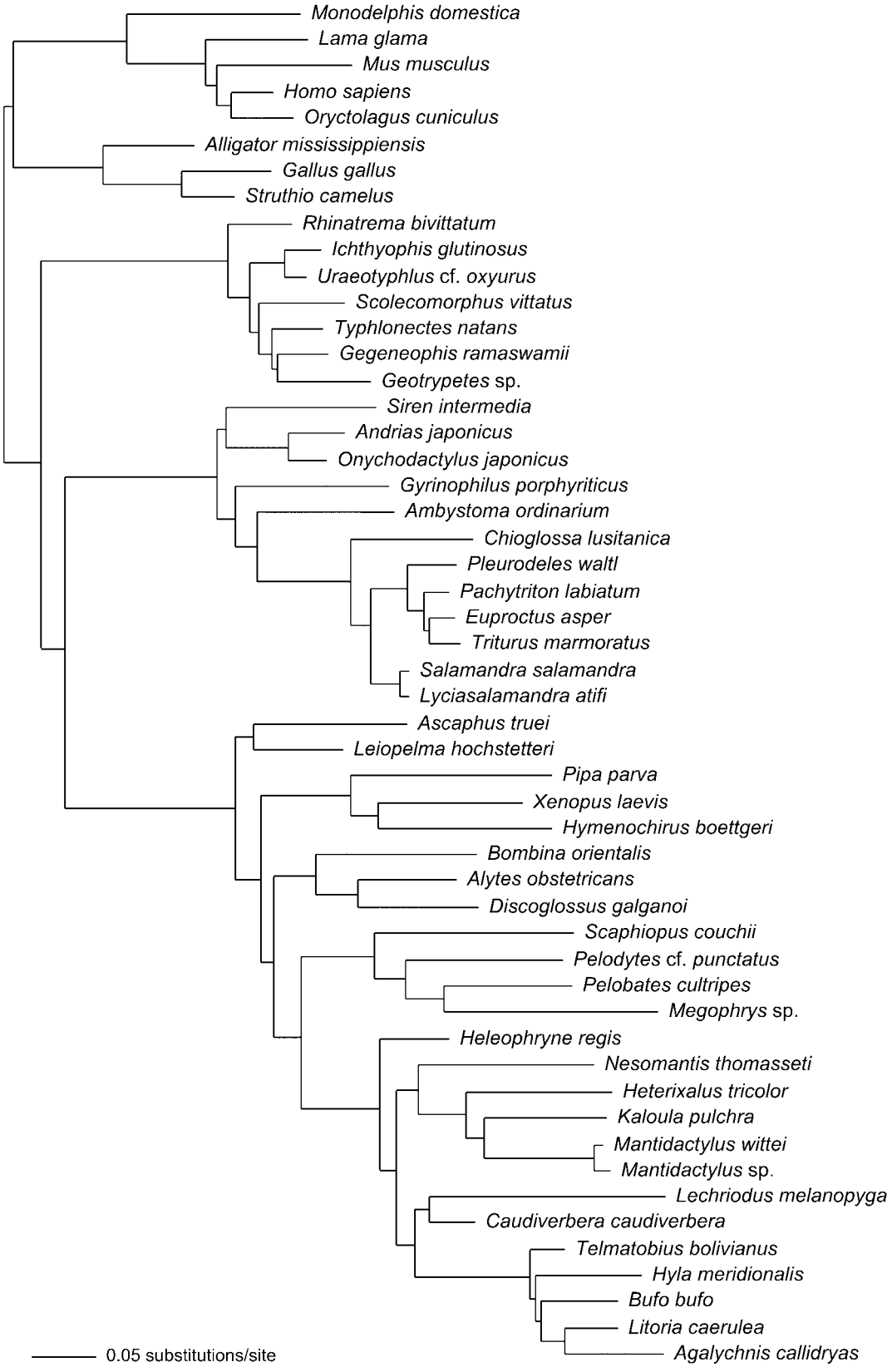


Figure C1: Unconstrained (nonultrametric) maximum likelihood phylogram showing the pattern of rate variability.

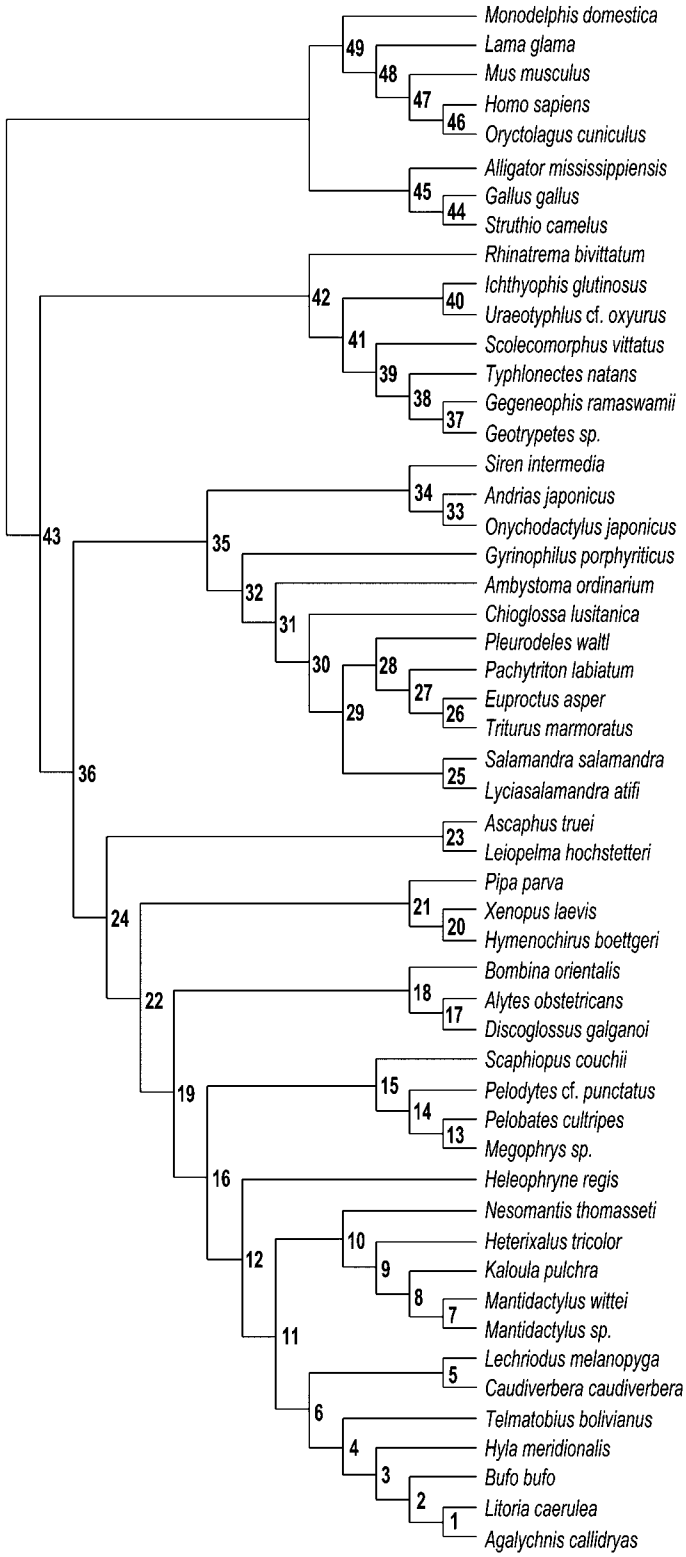


Figure C2: Topology of maximum likelihood tree with node numbers

Table C1
Statistical support and age estimates for each node

	Statistical support				Age estimate		
Node	ML	BI	ME	MP	Mean	SD	CI
1	99	100	62	78	46.786	4.507	42.136–58.676
2	…	…	…	…	57.932	6.894	46.703–73.632
3	62	80	…	…	60.911	7.313	48.793–77.276
4	100	100	100	100	65.054	7.976	51.916–82.813
5	59	96	…	…	120.302	16.153	90.608–154.232
6	…	66	…	…	138.029	16.259	108.041–171.613
7	100	100	100	100	9.783	2.835	4.455–14.604
8	77	85	54	75	86.527	14.961	58.430–117.456
9	100	100	98	93	99.044	15.831	69.940–132.137
10	51	69	…	…	130.791	17.253	99.065–166.582
11	…	99	…	…	149.698	17.569	117.136–186.236
12	100	100	98	95	161.706	18.087	127.892–198.934
13	77	100	57	51	117.860	21.874	77.179–161.823
14	55	97	…	…	142.478	22.350	100.511–186.596
15	100	100	100	97	163.582	22.470	121.398–208.485
16	75	100	…	…	216.353	21.210	176.105–259.721
17	100	100	96	95	151.529	24.047	104.840–199.422
18	100	100	100	97	198.574	22.812	155.298–243.508
19	58	96	…	…	234.862	21.336	194.878–277.135
20	54	93	75	60	154.815	25.044	107.122–204.529
21	100	100	100	98	176.697	24.524	130.300–226.269
22	82	98	…	…	244.773	21.597	204.091–288.157
23	86	91	100	94	236.661	22.575	192.567–281.017
24	100	100	100	100	262.470	20.798	223.183–304.551
25	100	100	100	100	35.001	15.688	11.878–72.953
26	61	90	66	66	37.216	12.877	17.492–67.106
27	100	100	96	99	45.100	14.773	22.066–78.803
28	100	100	97	100	66.955	18.681	36.038–108.754
29	89	98	100	91	109.771	22.505	68.634–155.305
30	100	100	100	100	138.300	24.024	92.913–186.583
31	82	100	…	…	230.107	22.234	187.516–274.276
32	63	98	…	…	252.585	20.505	213.251–293.835
33	100	100	100	100	177.404	13.696	161.513–211.731
34	…	…	…	…	261.358	18.722	226.966–300.312
35	100	100	100	100	272.544	19.036	238.099–311.837
36	…	67	71	75	356.959	22.261	317.256–405.261
37	65	95	…	…	146.417	14.070	130.529–182.564
38	58	99	…	89	155.237	15.275	134.243–193.482
39	56	99	78	88	177.064	17.788	147.982–217.600
40	100	100	98	100	104.305	22.436	64.645–151.029
41	56	94	83	94	192.414	18.786	160.280–232.994
42	100	100	100	100	214.285	20.306	177.412–255.705
43	94	100	76	71	367.378	22.699	327.517–417.364
44	100	100	100	100	120.214	28.004	70.487–177.954
45	100	100	100	100	237.672	24.284	185.812–279.520
46	71	83	78	91	116.492	26.867	66.555–170.896
47	80	85	…	65	142.652	28.232	87.831–196.801
48	100	100	100	100	159.483	28.201	103.183–213.141
49	100	100	100	100	236.372	24.267	183.620–278.913

Note: Statistical support given by bootstrap proportions for maximum likelihood (ML; 100 pseudoreplicates) and minimum evolution and maximum parsimony (ME and MP; 1,000 pseudoreplicates) and by Bayesian posterior probabilities (1,000,000 generations) for Bayesian inference (BI). Age estimates are in millions of years; also included are standard deviations and 95% confidence intervals (CI). Ellipses indicate support values of <50.

Appendix D from D. San Mauro et al., "Initial Diversification of Living Amphibians Predated the Breakup of Pangaea"

(Am. Nat., vol. 165, no. 5, p. 590)

Fit between Time and Divergence for the Employed Calibrations

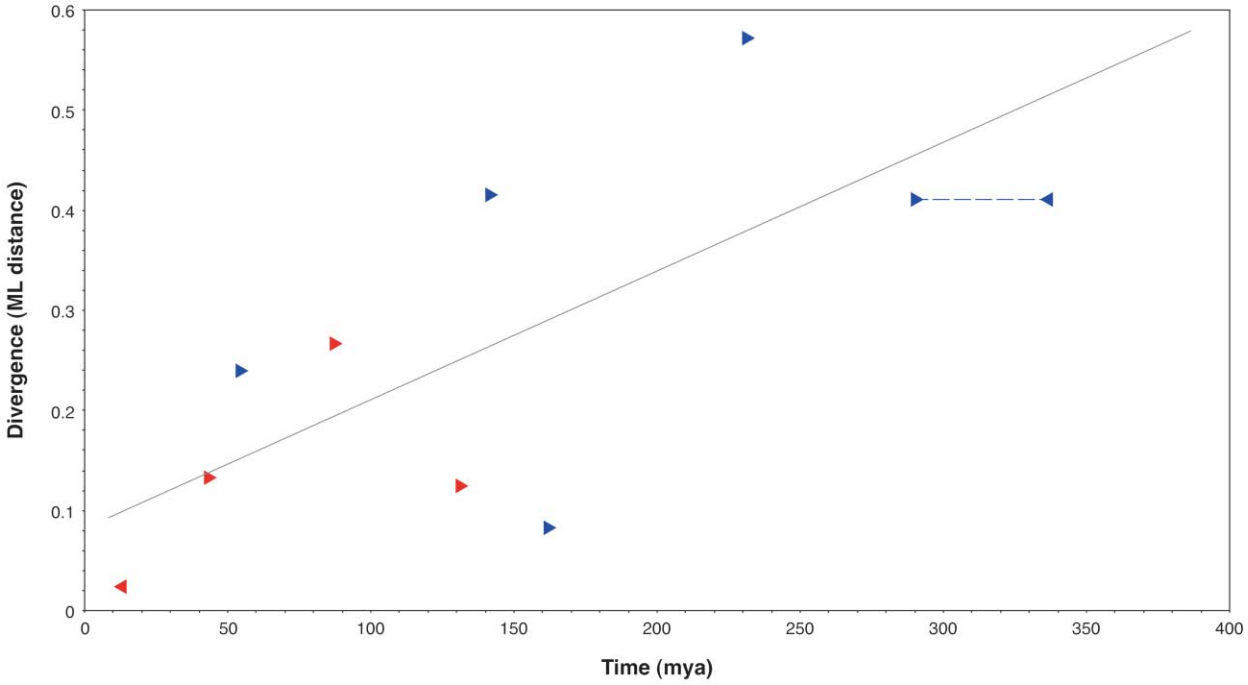


Figure D1: Scatterplot of divergence (measured as maximum likelihood [ML] distance) versus time (in millions of years) for the employed calibrations. These calibrations, as listed in "Material and Methods," are marked by triangles (upper and lower bounds). Red triangles represent calibrations based on biogeography; blue triangles represent calibrations based on fossil record. Dashed line indicates the interval for the synapsid-diapsid calibration. There is a significant correspondence between time and divergence even though the calibrations are not point calibrations but upper and lower time constraints (gray line represents linear regression fit; $R^2 = 0.458$; $F = 5.920$; $\text{df} = 1, 7$; $P = .045$).

4. PUBLICATION II

Title: **Phylogeny of caecilian amphibians (Gymnophiona) based on complete mitochondrial genomes and nuclear RAG1**

Authors: Diego San Mauro, David J. Gower, Oommen V. Oommen, Mark Wilkinson, Rafael Zardoya

Status: Published

Year: 2004

Journal: Molecular Phylogenetics and Evolution (SCI Impact Factor: 4.213)

Volume: 33

Pages: 413-427

Resumen II (Spanish translation of the abstract of Publication II)

Hemos determinado la secuencia nucleotídica completa del genoma mitocondrial (mt) de cinco cecilias (Amphibia: Gymnophiona) que representan cinco de las seis familias reconocidas: *Rhinatrema bivittatum* (Rhinatrematidae), *Ichthyophis glutinosus* (Ichthyophiidae), *Uraeotyphlus* cf. *oxyurus* (Uraeotyphlidae), *Scolecomorphus vittatus* (Scolecomorphidae), y *Gegeneophis ramaswamii* (Caeciliidae). La organización y tamaño de estos nuevos mitogenomas son similares a los anteriormente descritos para la cecilia *Typhlonectes natans* (Typhlonectidae), y para otros vertebrados. También se determinó la secuencia nucleotídica del gen nuclear RAG1 para estas seis especies de cecilias y la salamandra *Mertensiella luschani atifi*. El RAG1 (tanto a nivel de nucleótidos como de aminoácidos) muestra tasas de evolución más lentas que casi todos los genes mt codificantes de proteínas (a nivel de aminoácidos). Las nuevas secuencias mt y nucleares fueron comparadas con datos para otros anfibios y sometidas a análisis filogenéticos separados y combinados (Máxima Parsimonia, Mínima Evolución, Máxima Verosimilitud, e Inferencia Bayesiana). Todos los análisis apoyan con fuerza la monofilia de los tres órdenes de anfibios. La hipótesis Batrachia (Gymnophiona, (Anura, Caudata)) recibe apoyo moderado o bueno dependiendo del método de análisis. Dentro de Gymnophiona, el árbol óptimo (*Rhinatrema*, ((*Ichthyophis*, *Uraeotyphlus*), (*Scolecomorphus*, (*Gegeneophis*, *Typhlonectes*)))) es congruente con los estudios morfológicos y moleculares más recientes. La relación de grupo hermano entre Rhinatrematidae y todas las demás cecilias, entre Ichthyophiidae y Uraeotyphlidae, y la monofilia de las cecilias superiores Scolecomorphidae + Caeciliidae + Typhlonectidae, están todas fuertemente soportadas, mientras que las relaciones entre las cecilias superiores son resueltas de forma más ambigua. Los análisis del RAG1 están afectados por problemas de falso enraizamiento local y bajo soporte asociado que mejoran cuando los grupos externos son excluidos. La comparación de árboles usando los tests no paramétricos de Templeton, Kishino-Hasegawa, Aproximadamente Insesgado, y Shimodaira-Hasegawa sugieren que este último podría ser demasiado conservativo.

Available online at www.sciencedirect.com
SCIENCE DIRECT®
Molecular Phylogenetics and Evolution 33 (2004) 413–427
MOLECULAR PHYLOGENETICS AND EVOLUTION
www.elsevier.com/locate/ympev

Phylogeny of caecilian amphibians (Gymnophiona) based on complete mitochondrial genomes and nuclear RAG1

Diego San Mauro[a], David J. Gower[b], Oommen V. Oommen[c], Mark Wilkinson[b], Rafael Zardoya[a,*]

[a] *Departamento de Biodiversidad y Biología Evolutiva, Museo Nacional de Ciencias Naturales, CSIC, José Gutiérrez Abascal, 2, 28006 Madrid, Spain*
[b] *Department of Zoology, The Natural History Museum, Cromwell Road, London SW7 5BD, UK*
[c] *Department of Zoology, University of Kerala, Kariavattom 695 581, Thiruvananthapuram, Kerala, India*

Received 15 January 2004; revised 20 May 2004
Available online 28 July 2004

Abstract

We determined the complete nucleotide sequence of the mitochondrial (mt) genome of five individual caecilians (Amphibia: Gymnophiona) representing five of the six recognized families: *Rhinatrema bivittatum* (Rhinatrematidae), *Ichthyophis glutinosus* (Ichthyophiidae), *Uraeotyphlus* cf. *oxyurus* (Uraeotyphlidae), *Scolecomorphus vittatus* (Scolecomorphidae), and *Gegeneophis ramaswamii* (Caeciliidae). The organization and size of these newly determined mitogenomes are similar to those previously reported for the caecilian *Typhlonectes natans* (Typhlonectidae), and for other vertebrates. Nucleotide sequences of the nuclear RAG1 gene were also determined for these six species of caecilians, and the salamander *Mertensiella luschani atifi*. RAG1 (both at the amino acid and nucleotide level) shows slower rates of evolution than almost all mt protein-coding genes (at the amino acid level). The new mt and nuclear sequences were compared with data for other amphibians and subjected to separate and combined phylogenetic analyses (Maximum Parsimony, Minimum Evolution, Maximum Likelihood, and Bayesian Inference). All analyses strongly support the monophyly of the three amphibian Orders. The Batrachia hypothesis (Gymnophiona, (Anura, Caudata)) receives moderate or good support depending on the method of analysis. Within Gymnophiona, the optimal tree (*Rhinatrema*, ((*Ichthyophis*, *Uraeotyphlus*), (*Scolecomorphus*, (*Gegeneophis Typhlonectes*)))) agrees with the most recent morphological and molecular studies. The sister group relationship between Rhinatrematidae and all other caecilians, that between Ichthyophiidae and Uraeotyphlidae, and the monophyly of the higher caecilians Scolecomorphidae + Caeciliidae + Typhlonectidae, are strongly supported, whereas the relationships among the higher caecilians are less unambiguously resolved. Analysis of RAG1 is affected by a spurious local rooting problem and associated low support that is ameliorated when outgroups are excluded. Comparisons of trees using the non-parametric Templeton, Kishino–Hasegawa, Approximately Unbiased, and Shimodaira–Hasegawa tests suggest that the latter may be too conservative.

Keywords: Gymnophiona; Amphibia; Molecular phylogenetics; Mitochondrial genome; RAG1

1. Introduction

Caecilians (Gymnophiona) are one of the three orders of recent Amphibia. They are readily distinguished from frogs and salamanders by their sensory tentacles and annulated, limbless bodies, and are distinct in many other characters (e.g., Himstedt, 1996; Noble, 1931; Taylor, 1968). Most of the approximately 160 currently recognized species (Frost, 2002; Nussbaum and Wilkinson, 1989) are tropical, soil-dwelling predators for at least their adulthood, but members of the South American family Typhlonectidae are semiaquatic or aquatic (Wilkinson and Nussbaum, 1999). Despite increasing evidence of high local abundance in some species (e.g., Gower et al., 2004; Measey et al., 2003), caecilians

* Corresponding author. Fax: +34-91-564-5078.
E-mail address: rafaz@mncn.csic.es (R. Zardoya).

1055-7903/$ - see front matter © 2004 Elsevier Inc. All rights reserved.
doi:10.1016/j.ympev.2004.05.014

remain probably the least known order of recent tetrapods. Although they are a relatively small group, it is clear that they have a remarkable morphological, ecological, and reproductive diversity (e.g., Gower et al., 2004; Loader et al., 2003; Wake, 1977; Wilkinson and Nussbaum, 1997).

Fossils of potential crown-group caecilians are represented only by isolated vertebrae (Evans and Sigogneau-Russel, 2001) so that inferring phylogenetic relationships among caecilians is essentially a neontological enterprise. Between 1968 and 1979, the previously single caecilian family Caeciliidae was partitioned into the six higher taxa (Nussbaum, 1977, 1979; Taylor, 1968, 1969) that are widely recognized as families today (Duellman and Trueb, 1986; Nussbaum and Wilkinson, 1989; Wilkinson and Nussbaum, 1999). Four caecilian families have relatively restricted distributions—the South American Rhinatrematidae (two genera, eight species) and Typhlonectidae (five genera, 13 species), Indian Uraeotyphlidae (one genus, five species), and African Scolecomorphidae (two genera, six species). Ichthyophiidae (two genera, 30+ species) occurs in South and South East Asia (West of Wallace's line). The more cosmopolitan Caeciliidae (21 genera, ca. 100 species) occurs on all land masses where caecilians are known except South East Asia. The current distribution of extant caecilians is consistent with a Gondwanan origin of the order (Duellman and Trueb, 1986; Hedges et al., 1993; Wilkinson et al., 2002). The four smaller and more local families represent morphologically distinctive caecilian clades (Nussbaum, 1977, 1979, 1985; Wilkinson and Nussbaum, 1999). In contrast, molecular data suggest that ichthyophiids might not be monophyletic (Gower et al., 2002), and morphology and molecules agree that the Caeciliidae, which comprises those caecilians that have not been assigned to the five more recently described families, is paraphyletic (Hedges et al., 1993; Nussbaum, 1979; Wilkinson, 1997; Wilkinson et al., 2003).

Nussbaum (1979) presented the first numerical phylogenetic analysis of caecilians, using morphological characters to investigate the interrelationships of 12 genera. This, and the subsequent analyses of Duellman and Trueb (1986) and Hillis (1991) that used family-level taxa and a subset of Nussbaum's (1979) characters, identified a clade comprising the caeciliids, typhlonectids, and scolecomorphids that Nussbaum (1991) dubbed the "higher" caecilians. The Uraeotyphlidae, Ichthyophiidae, and Rhinatrematidae were successively more distant outgroups to the higher caecilians in these analyses. Diverse morphological evidence that the Rhinatrematidae is the sister group of all other extant caecilians (Nussbaum, 1977; Wilkinson, 1992, 1996a) is considered to provide strong support for this hypothesis, which has been used to root caecilian phylogenetic trees in more recent morphological and molecular analyses (Gower et al., 2002; Wilkinson, 1997; Wilkinson and Nussbaum, 1996; Wilkinson et al., 2002, 2003). Wilkinson and Nussbaum (1996) and Wilkinson (1997) also supported the monophyly of the higher caecilians, but found strong support for an alternative arrangement of more deep-branching families, in which the Ichthyophiidae and Uraeotyphlidae are sister taxa. Whereas earlier family-level phylogenies (Duellman and Trueb, 1986; Hillis, 1991) recovered Caeciliidae and Typhlonectidae as more closely related to each other than to Scolecomorphidae, the most comprehensive morphological study to date (Wilkinson, 1997) was unable to resolve relationships among these higher caecilians.

Previous molecular analyses that have been informative regarding the relationships among caecilian families have used nucleotide sequences of mitochondrial (mt) cytochrome *b* and 12S and 16S rRNA genes (Gower et al., 2002; Hay et al., 1995; Hedges and Maxson, 1993; Hedges et al., 1993; Wilkinson et al., 2002, 2003). These have supported recent morphological analyses by recovering clades comprising Ichthyophiidae + Uraeotyphlidae and Nussbaum's (1991) higher caecilians (caeciliids, scolecomorphids, and typhlonectids) (Wilkinson et al., 2003), and a paraphyletic Caeciliidae (Hedges and Maxson, 1993; Hedges et al., 1993; Wilkinson et al., 2002, 2003). Wilkinson et al. (2003) carried out the only previous molecular analysis to include members of all six currently recognized families. In agreement with the most recent morphological investigation, their study suggested that Caeciliidae is paraphyletic with respect to perhaps Scolecomorphidae as well as Typhlonectidae. However, many relationships within the higher caecilians were not strongly supported, and they suggested that more molecular and morphological data were required to resolve these relationships.

We have determined the complete nucleotide sequences of the mt genomes of five caecilian species, and compared them with the only previously described caecilian mt genome, that of *Typhlonectes natans* (Zardoya and Meyer, 2000). The sampling includes one representative of each of the six currently recognized families. Our mitogenomic (Curole and Kocher, 1999) approach follows several recent studies (Cummings et al., 1995; Russo et al., 1996; Zardoya and Meyer, 1996b) that demonstrated the utility of large sequence data sets for establishing robust high-level phylogenetic inferences. To provide independent data from a different genome, we have also sequenced the nuclear gene RAG1, which has proven useful in inferring relationships among other major vertebrate lineages (e.g., Groth and Barrowclough, 1999; Martin, 1999). Through separate and combined analyses, we explore the utility of these data in establishing a robust higher-level phylogenetic framework for caecilians. The inclusion of comparable data for representatives of frogs, salamanders, and more dis-

tant outgroups, make our analyses relevant to the controversial relationships (Zardoya and Meyer, 2001) among the three recent amphibian orders.

2. Materials and methods

2.1. Taxon sampling

Our sampling includes one species from each of the six currently recognized caecilian families. The typhlonectid *T. natans* was chosen because of the availability of its mt genome sequence (Zardoya and Meyer, 2000). The other five species belong to the type genus (sometimes as the type species) of their respective families (Table 1). Caeciliid paraphyly means that this family is inadequately represented with a single species. The type genus, the Neotropical *Caecilia*, has been shown to be among those caeciliids most closely related to Typhlonectidae (Hedges et al., 1993; Wilkinson et al., 2002, 2003). In contrast, our chosen caeciliid, the Indian *Gegeneophis ramaswamii*, is more distantly related to the Neotropical typhlonectids than is *Caecilia* (Wilkinson et al., 2002), thus providing an opportunity to further explore the nature of caeciliid paraphyly.

For comparisons of mt genomes, we selected the following outgroups (GenBank accession numbers in parentheses): the anuran amphibians *Xenopus laevis* (NC_001573, Roe et al., 1985) and *Rana nigromaculata* (NC_002805, Sumida et al., 2001), the caudate amphibians *Mertensiella luschani* (NC_002756, Zardoya et al., 2003) and *Ranodon sibiricus* (NC_004021, Zhang et al., 2003), and two lobe-finned fishes, a coelacanth, *Latimeria chalumnae* (NC_001804, Zardoya and Meyer, 1997), and an African lungfish, *Protopterus dolloi* (NC_001708, Zardoya and Meyer, 1996a).

Outgroups for examination of RAG1 sequences were one anuran, *X. laevis* (L19324, Greenhalgh et al., 1993), two caudates, *Pleurodeles waltl* (AJ010258, Frippiat et al., 2001) and *Mertensiella luschani atifi* (Table 1), a coelacanth, *Latimeria menadoensis* (AY442925, Brinkmann et al., 2004); and an African lungfish, *P. dolloi* (AY442928, Brinkmann et al., 2004).

2.2. DNA extraction, PCR amplification, cloning, and sequencing

Total DNA was purified from ethanol-preserved liver or muscle, with a standard phenol/chloroform extraction procedure (Sambrook et al., 1989). A suite of 28 primers (Table 2) was used to amplify by PCR contiguous and overlapping fragments that covered the entire mt genome (Fig. 1). PCR amplifications were conducted in 25 μl reactions containing 67 mM Tris–HCl, pH 8.3, 1.5 mM $MgCl_2$, 0.4 mM of each dNTP, 2.5 μM of each primer, template mtDNA (10–100 ng), and *Taq* DNA polymerase (1 U, Biotools), using the following cycling conditions: an initial denaturing step at 94 °C for 5 min; 35 cycles of denaturing at 94 °C for 60 s, annealing at 42–54 °C (see Table 2) for 60 s, and extending at 72 °C for 90 s; and a final extending step of 72 °C for 7 min. PCR products were purified by ethanol precipitation, and sequenced in an automated DNA sequencer (ABI PRISM 3700) using the BigDye Deoxy Terminator cycle-sequencing kit (Applied Biosystems) following manufacturer's instructions. For mtDNA, short amplicons were sequenced directly using the corresponding PCR primers. Long amplicons were cloned into pGEM-T vectors (Promega), and recombinant plasmids were sequenced using the M13 (forward and reverse) universal primers as well as additional walking primers (available from the authors upon request). The sequences obtained averaged 700 base pairs (bp) in length, and each sequence overlapped the next contig by about 150 bp. In no case were differences in sequence observed between the overlapping regions.

Four primers were designed in conserved regions of the RAG1 gene to amplify, by PCR, two contiguous

Table 1
Data for amphibian samples employed in this study

Species	Taxonomic assignment	Voucher No.	Collection locality	GenBank Accession Nos. (mt genomes, RAG1)
Gegeneophis ramaswamii	Gymnophiona: Caeciliidae	MW 331	Thenmalai, India	AY456250, AY456255
Ichthyophis glutinosus[c]	Gymnophiona: Ichthyophiidae	MW 1733	Peradeniya, Sri Lanka	AY456251, AY456256
Rhinatrema bivittatum[c]	Gymnophiona: Rhinatrematidae	BMNH 2002.6	Kaw, French Guyana	AY456252, AY456257
Scolecomorphus vittatus[c]	Gymnophiona: Scolecomorphidae	BMNH 2002.100	Amani, Tanzania	AY456253, AY456258
Uraeotyphlus cf. *oxyurus*[c]	Gymnophiona: Uraeotyphlidae	MW 212	Payyanur, India	AY456254, AY456259
Typhlonectes[c] *natans*	Gymnophiona: Typhlonectidae	BMNH 2000.218	Potrerito, Venezuela	AF154051,[a] AY456260
Mertensiella luschani atifi	Caudata: Salamandridae	—	Fersin, Turkey	NC_002756,[b] AY456261

BMNH, The Natural History Museum, London; MW, field series of the Zoology Department, University of Kerala (India); and the Department of National Museums, Colombo (Sri Lanka).

[a] Zardoya and Meyer (2000).

[b] Zardoya et al. (2003).

[c] Indicates type species/genus of family.

Table 2
Primers used to sequence the complete caecilian mt genomes (see Fig. 1 to trace fragments along the genome)

Fragment name	Primer name	Sequence	Approximate product length (bp)	Annealing temperature (°C) used in the PCR
12S	L1091[a]	5′-AAAAAGCTTCAAACTGGGATTAGATACCCCACTAT-3′	380	51
	H1478[a]	5′-TGACTGCAGAGGGTGACGGGCGGTGTGT-3′		
MID	Amp-12S F	5′-AAGAAATGGGCTACATTTTCT-3′	1200	50
	Amp-16S R	5′-AAGTGATTAYGCTACCTTTGCAC-3′		
16S	16Sar-L[b]	5′-CGCCTGTTTATCAAAAACAT-3′	500	51
	16Sbr-H[b]	5′-CCGGTCTGAACTCAGATCACGT-3′		
P1	MNCN-16S F[c]	5′-GGTTTACGACCTCGATGTTGGATC-3′	1350	42
	Lati-Met R[d]	5′-TCGGGGTATGGGCCCGAAAGCTT-3′		
P2	Amp-P2 F	5′-CAAYTAATRCAYCTAGTATGRAAAA-3′	2500	42
	Amp-P2 R	5′-ATATARCCAAAWGGTTCTTTTTT-3′		
P3	Amp-P3 F	5′-CAATACCAAACCCCCTTRTTYGTWTGATC-3′	900	45
	Amp-P3 R	5′-GCTTCTCARATAATAAATATYAT-3′		
P4	Amp-P4 F	5′-GGMTTTATTCACTGATTYCC-3′	1400	50
	Amp-P4 R	5′-AAATTGGTCAAAKAARCTTAGKRTCATGGTCA-3′		
P5	8.2 L8331[e]	5′-AAAGCRTYRGCCTTTTAAGC-3′	1590	54
	MNCN-COIII R[c]	5′-ACGTCTACRAARTGTCAGTATCA-3′		
P6	Amp-P6 F	5′-ACATGAGCYCAYCACAGYATTAT-3′	1440	50
	Amp-P6 R	5′-CGGGTAATAATAATTAATGTTGG-3′		
P7	Amp-P7 F	5′-AAYCTCCTACAATGYTAAAAAT-3′	1550	48
	Amp-P7 R	5′-CATARCTTTTACATGGATTTGCACC-3′		
P8	MNCN-His F[c]	5′-AAAACATTAGATTGTGATTCTAA-3′	1210	42
	Lati-ND5 R1[d]	5′-CCYATYTTTCKGATRTCYTGYTC-3′		
P9	Amp-P9 F	5′-AGCCARCTYGGCCTAATAATAGT-3′	1630	50
	Amp-P9 R	5′-CAGCCGTARTTTACGTCTCGRCAGAT-3′		
P10	MNCN-Glu F[c]	5′-GAAAAACCACCGTTGTTATTCAACTACA-3′	1170	48
	Amp-P10 R	5′-TTCAGYTTACAAGACYGATGCTTT-3′		
P11	Amp-P11 F	5′-TGRATYGGRGGCCAACCAGTAGAAGA-3′	1550	50
	Amp-12S R	5′-TCGATTATAGAACAGGCTCCTCT-3′		

[a] Kocher et al. (1989).
[b] Palumbi et al. (1991).
[c] Zardoya (Unpublished data).
[d] Zardoya and Meyer (1997).
[e] http://nmg.si.edu/bermlab.htm.

and overlapping fragments that cover a 1500 bp portion of the 3′ end of this gene: Amp-RAG1 F (5′-AGC TGC AGY CAR TAC CAY AAR ATG TA-3′), Amp-RAG1 R1 (5′-AAC TCA GCT GCA TTK CCA ATR TCA CA-3′), Amp-RAG1 F1 (5′-ACA GGA TAT GAT GAR AAG CTT GT-3′), and Amp-RAG1 R (5′-TTR GAT GTG TAG AGC CAG TGG TGY TT-3′). PCR mixtures and cycling conditions were as described above (annealing temperature was 54 °C). PCR products were cloned into pGEM-T vectors and sequenced using the M13 universal primers as described above.

All new nucleotide sequences reported in this paper have been deposited in GenBank under accession numbers given in Table 1.

2.3. Molecular and phylogenetic analyses

Nucleotide sequences (RAG1 gene) and the deduced amino acid sequences of all 13 mt protein-coding genes were aligned separately, using the default parameters of CLUSTAL X version 1.83 (Thompson et al., 1997), and the alignments revised by eye in an effort to maximize

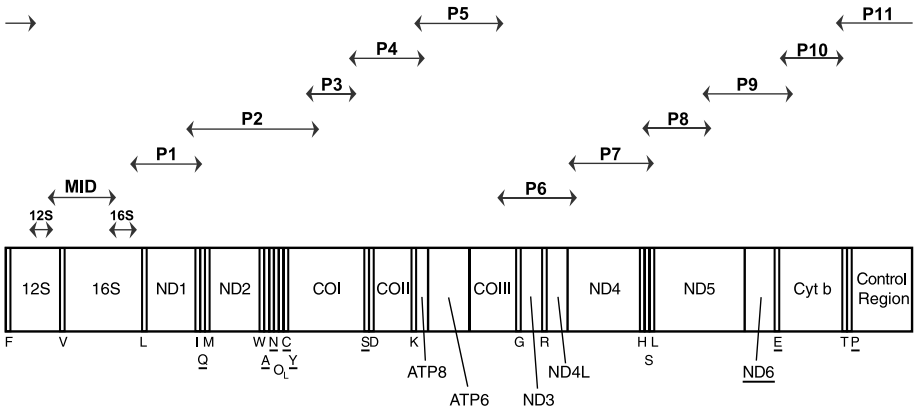


Fig. 1. Gene organization and sequencing strategy for the mt genomes of the caecilians. Genes encoded by the L strand are underlined. Arrow-headed segments denote the location of the fragments amplified by PCR with each pair of primers (see Table 2 for the primer DNA sequence associated with each fragment). *Gegeneophis ramaswamii* departs from this general consensus in lacking the $tRNA^{Phe}$ gene.

positional homology. Alignment ambiguities and gaps were excluded from phylogenetic analyses using GBLOCKS version 0.91b (Castresana, 2000). Alignments and data files are available from the authors upon request.

Four commonly used methods of phylogenetic inference, namely Maximum Parsimony -MP- (Fitch, 1971), Minimum Evolution -ME- (Rzhetsky and Nei, 1992), Maximum Likelihood -ML- (Felsenstein, 1981), and Bayesian Inference -BI- (Huelsenbeck et al., 2001) were applied separately to the RAG1 data and to a concatenated dataset of the amino acid sequences of the mt protein-coding genes. Bayesian and MP analyses of the combined RAG1 (except *Pleurodeles*) and mt amino acid sequence data (except *Ranodon* and *Rana*) were performed. Separate analyses using only the ingroup taxa (caecilian-only data) were also performed for both mt amino acid and RAG1 nucleotide data sets.

Quartet puzzling ML analyses of amino acid sequence data (100,000 puzzling steps) were conducted with TREE-PUZZLE version 5.0 (Strimmer and von Haeseler, 1996). ML analysis of RAG1, and all ME and MP analyses, were performed with PAUP* version 4.0b10 (Swofford, 1998), with 10 random addition sequences and TBR branch swapping. ME analyses of mt amino acid and nuclear DNA sequences used mean character and ML distances, respectively. BIs were made using MrBayes version 3.0b4 (Huelsenbeck and Ronquist, 2001) with four simultaneous chains, each of a million generations, sampled every 100 generations. Trees sampled before the cold chain reached stationarity, as judged by plots of ML scores, were discarded as "burn-in."

Following Yang et al. (1998), we used the mtREV24 model (Adachi and Hasegawa, 1996) in all likelihood and Bayesian analyses of amino acid data, and we employed likelihood ratio tests (LRTs) to select among the following hierarchically nested alternative models: equal rates (eq.) versus proportion of invariant sites (I), versus gamma-distributed rates (Γ), versus gamma-distributed rates and proportion of invariant sites (Γ+I). ML analyses of RAG1 sequences used the best-fit model of nucleotide substitution selected according to the Akaike information criterion (AIC) calculated using Modeltest version 3.4 (Posada and Crandall, 1998). For BI, best-fit models were selected for each RAG1 codon position and model parameters were estimated independently ("unlink" option). For the combined data Bayesian analysis, and for analyses of the caecilian-only data, best-fit models were re-estimated for each partition because of the exclusion of taxa.

Support was evaluated with non-parametric bootstrap proportions (BPs—1000 pseudoreplicates), Bayesian posterior probabilities (BPPs), and quartet puzzling proportions (QPs). Decay indices (*d*) were also calculated using AutoDecay version 5.04 (Eriksson, 2001). Approximately Unbiased -AU- (Shimodaira, 2002), Shimodaira–Hasegawa -SH- (Shimodaira and Hasegawa, 1999), Kishino–Hasegawa -KH- (Kishino and Hasegawa, 1989), and Templeton (Templeton, 1983) tests were used to evaluate the 105 alternative, fully resolved unrooted trees for the caecilian-only data. Templeton test (two tailed) was performed in PAUP*, whereas the other three tests were carried out using CONSEL version 0.1f (Shimodaira and Hasegawa, 2001) with site likelihoods calculated by p4 version 0.79 (Foster, 2003).

Substitution rates and among-site rate heterogeneities were compared among RAG1 (at both nucleotide and amino acid levels), each mt protein (at the amino acid level), and a concatenated data set including all mt proteins (at the amino acid level) using the same subset of taxa used in the combined BI (see above). BI (100,000 generations) was used to estimate substitution rate (measured as tree length -TL-) and among-site rate heterogeneity (α parameter of the gamma distribution).

ANOVA analyses were used to assess variations in substitution rates and among-site rate heterogeneities. Planned comparisons were used to examine contrasts between RAG1 and each mt protein, and between RAG1 and the concatenated mt data set. Statistical analyses were performed using STATISTICA version 6.0 (StatSoft Inc., 2001).

3. Results

3.1. Mitochondrial genome organization and structural features

The complete nucleotide sequences of the L strands of the mt genomes of five caecilians (*G. ramaswamii*, *Ichthyophis glutinosus*, *Rhinatrema bivittatum*, *Scolecomorphus vittatus*, and *Uraeotyphlus* cf. *oxyurus*) were determined. Total length ranged from 15,897 to 16,432 bp. As in *T. natans*, all five newly sequenced caecilian mt genomes encoded for two rRNAs, 22 tRNAs, and 13 protein-coding genes, with the single exception of *G. ramaswamii*'s lack of the $tRNA^{Phe}$ gene. In all cases, the organization (Fig. 1) conforms to the consensus mt gene arrangement for vertebrates (Jameson et al., 2003). Other notable distinct features are only found in non-coding regions.

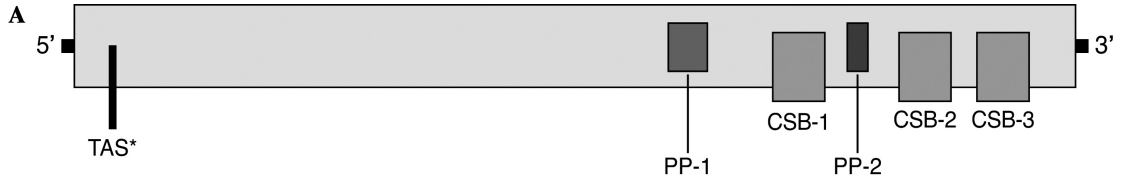
The control regions of the five new caecilian mt genomes are similar in length, ranging from 600 to 682 bp, and are also similar in structure and motifs (Fig. 2A). Three conserved blocks (CSB-1, CSB-2, and CSB-3, Walberg and Clayton, 1981) were identified at the 3′ end of each control region (Fig. 2B). Two polypyrimidine tracts, PP-1 and PP-2, were identified upstream from the CSB-2 and CSB-3 motifs (Fig. 2A). PP-1 consists of a stretch of thymines, and PP-2 is a poly(C) stretch located between CSB-1 and CSB-2 motifs. A putative termination-associated sequence (TAS) was found only in *S. vittatus*, close to the 5′ end of the control region. In contrast to *T. natans* (Zardoya and Meyer, 2000), no tandem repeats were found in the control regions of the newly sequenced caecilian mt genomes.

As in most vertebrates, the putative origin of L-strand replication (O_L) of the five new caecilian mt genomes was located within the WANCY tRNA cluster, between the $tRNA^{Asn}$ and $tRNA^{Cys}$ genes (Fig. 1). The O_L ranges from 30 to 39 bp and, in all five caecilians, has the potential to fold into a stem–loop secondary structure, sharing some nucleotides with the flanking tRNAs (Fig. 3). However, none of them can fold into alternative secondary structures with the adjacent $tRNA^{Cys}$ sequence such as reported for *T. natans* (Zardoya and Meyer, 2000). The 5′-GCCGG-3′ motif that in human mtDNA is involved in the transition from RNA synthesis to DNA synthesis (Hixson et al., 1986), is entirely conserved in the mtDNA of *R. bivittatum* and *U.* cf. *oxyurus*, whereas the remaining caecilian mt genomes show less conserved motifs (Fig. 3).

The mt genomes of *R. bivittatum* and *U.* cf. *oxyurus* have long non-coding regions between $tRNA^{Thr}$ and $tRNA^{Pro}$ genes of 312 and 437 bp, respectively. No secondary structures, tandem repeats, or functional ORFs were found in these intergenic regions, and BLAST searches produced no close matches. The non-coding spacer of *R. bivittatum* exhibits the same base composition as in the L strand of the whole mt genome, whereas in *U.* cf. *oxyurus* there is a much higher frequency of C (33%) and lower frequency of G (8%) than in the L strand.

3.2. RAG1 molecular features

All amphibian RAG1 sequences examined in this study are very conserved and show no indels. Absence of a single codon distinguishes amphibians from lobe-finned fishes. Overall base frequencies of sequences of



B

	CSB-1	CSB-2	CSB-3
Gegeneophis ramaswamii	TATATAAT-GAATGCTTGACGGACATA	TAAACCCCCC-TACCCCCC	CGTCAAACCCCAAAATCCA
Ichthyophis glutinosus	TAATAAAT-GAATGACTTATGGACATA	TAAACCCCCCTTACCCCCC	CGTCAAACCCCAAAATCCG
Rhinatrema bivittatum	T-ATATAT-GAATGATTTATTGACATA	CAAACCCCCCCT-CCCCCC	CGTCAAACCCCTAAATCCG
Scolecomorphus vittatus	TAATAAAT-GAATGTATTGAGGACATA	TAAACCCCCCTTACCCCCC	AGCTAAACCCCTAAATCTG
Uraeotyphlus cf. *oxyurus*	TTATAAAT-GAATGATTTATGGACATA	CAAACCCCCCCTACCCCCC	CGTCAAACCCCAAAATCCG
Typhlonectes natans	TCATAAATTGAATGTTTAATTGATACA	TAAACCCCCC-T--CCCCC	CGCTAAACCCCTAAATTCA

Fig. 2. Main features of the caecilian mtDNA control region. (A) Consensus structure of the control region. All caecilians have three conserved sequence blocks (CSB-1, 2, and 3) and two pyrimidine-rich regions (PP-1 and 2). TAS* is found only in *Scolecomorphus vittatus* and *Typhlonectes natans*. The latter taxon possesses, in addition, seven 109-bp tandem repeats in the right domain, close to the 3′ end. (B) Alignments of the identified CSBs in caecilians. Data for *Typhlonectes natans* are from Zardoya and Meyer (2000).

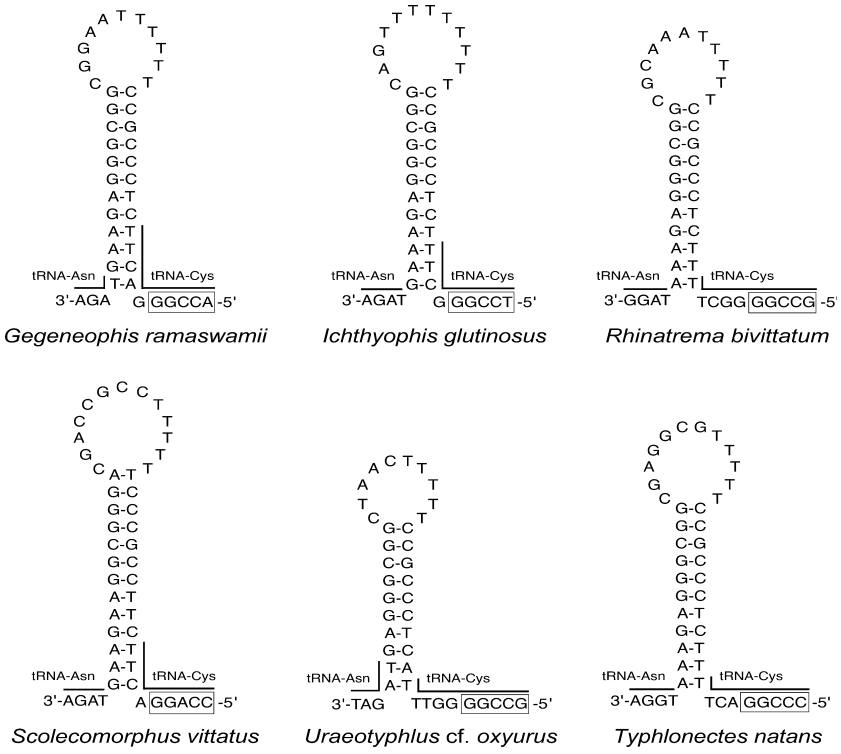


Fig. 3. Proposed secondary structures for the origins of L-strand replication (O_L) in caecilians. The 5′-GCCGG-3′ related motif is indicated by a box. Lines show the nucleotides partially shared with flanking tRNAs. Data for *Typhlonectes natans* are from Zardoya and Meyer (2000).

the frog and salamanders are fairly similar, but sequences of caecilians and lobe-finned fishes have a higher frequency of A (30.2–32.5%) and lower frequency of C (17.7–20.0%). Third codon positions in frogs, caecilians, and lobe-finned fishes have a G+C content of 31.8–47.6%, whereas the salamanders have a moderately strong G+C bias (61.0–62.8%), a difference that is significant in pairwise comparisons ($\chi^2_{5\%(30)} = 202.49$, $P < 0.001$). Pairwise differences among other codon positions across all taxa are not significant.

Average substitution rate of RAG1 is relatively low in comparison with those of most mt proteins (Fig. 4). RAG1 rates (at the nucleotide level) are between those of COII and COIII (at the amino acid level), whereas RAG1 rates at the amino acid level are between rates of COIII and COI (at the amino acid level) (Fig. 4). Relative substitution rates (TL) estimated for the different mt proteins are highly variable (Fig. 4). All statistical contrasts between RAG1 and each mt protein, and between RAG1 and the concatenated mt data set are highly significant ($F_{1,14400}$ values range from 17.79 to 75,809.88; $P < 0.001$ in all cases). For the caecilian-only data, all mt proteins (and the concatenated mt data set) have TL values about half those for the all-taxa data, whereas for RAG1 the differences in TL values are much greater (about 4.6 times at the nucleotide level and 3.6 times at the amino acid level).

Estimated among-site rate heterogeneities (α) are quite similar among the different protein data sets (Fig. 4). With the exception of ATP8 and ND3, α is less than two for all subsets of mt data. RAG1 among-site rate heterogeneity at the nucleotide level is 3.32 ± 0.09, lying between the values for COIII and ND3 (at the amino acid level). RAG1 among-site variation at the amino acid level is 1.17 ± 0.03, lying between the values for ND4 and ND4L (at the amino acid level) (Fig. 4). All statistical contrasts between α values for RAG1 at the nucleotide level and each mt data set are highly significant ($F_{1,14400}$ values ranged from 41.49 to 2579.20; $P < 0.001$ in all cases). At the amino acid level, only contrasts between RAG1 and ATP8, COIII, ND2, ND3, and ND6 are significant ($F_{1,14400}$ values ranged from 5.26 to 3826.08; $P < 0.05$ in all five cases).

3.3. Phylogenetic analyses

The deduced amino acid sequences of all 13 mt protein-coding genes of six caecilians, two salamanders, two frogs, and two lobe-finned fishes were combined into a single data set that produced an alignment of 3857 positions. Of these, 394 were excluded from the analyses because of alignment ambiguities, 1615 are invariant, and 1179 are parsimony informative. Within caecilians, the number of parsimony-informative sites is 518. Mean character distances among caecilians range from 0.15 (*Ichthyophis* vs. *Uraeotyphlus*) to 0.25 (*Ichthyophis* vs. *Typhlonectes*), and among amphibian orders from 0.23 to 0.32. MtREV24+Γ+I was selected as the best-fitting model.

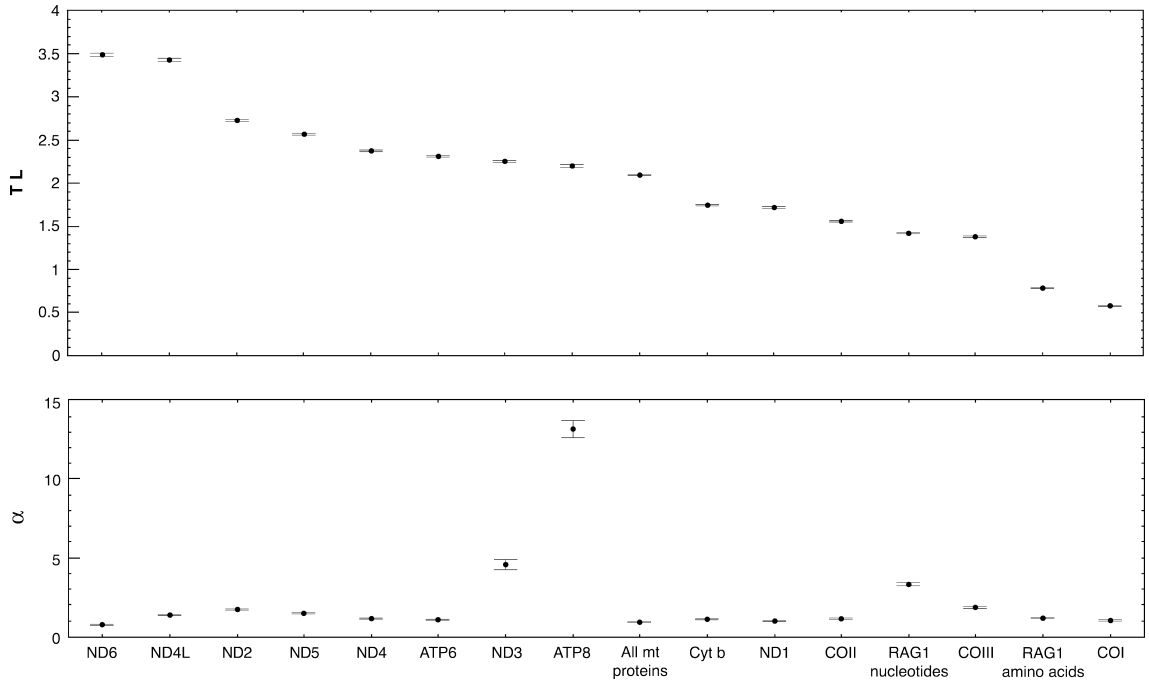


Fig. 4. Substitution rates (measured as Bayesian tree length -TL-) and among-site rate heterogeneities (α parameter of the gamma distribution) of RAG1 (at both nucleotide and amino acid levels), each mt protein (at the amino acid level), and a concatenated data set including all mt proteins (at the amino acid level).

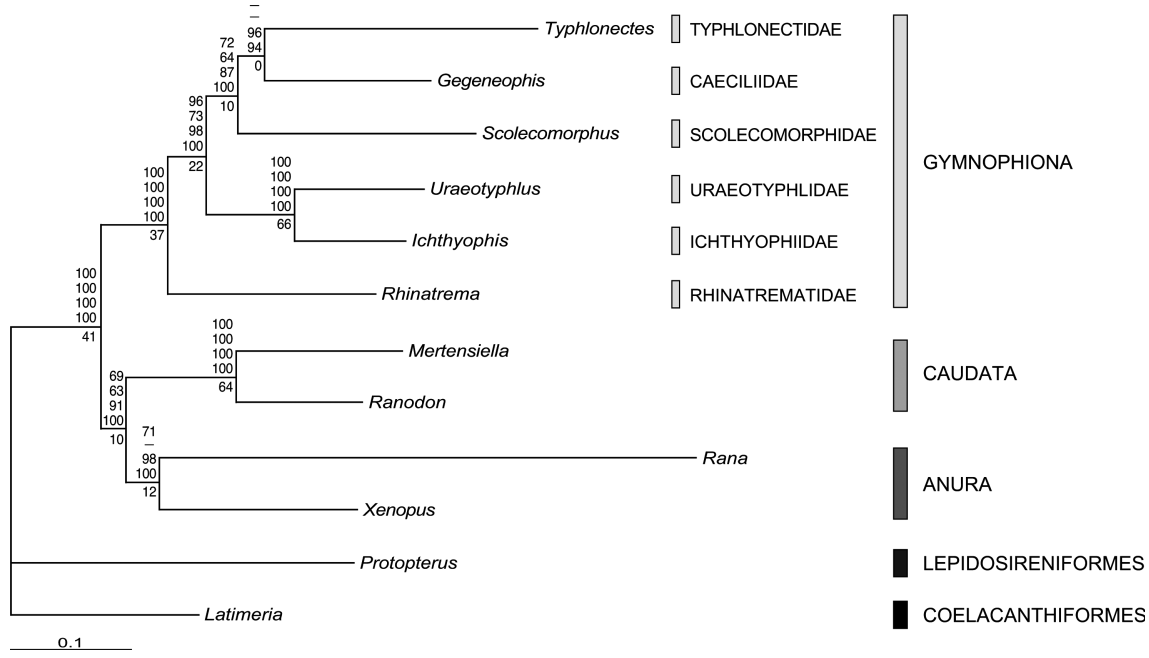


Fig. 5. Phylogenetic relationships (ML phylogram) of caecilians inferred from a single concatenated data set of the deduced amino acid sequences of all 13 mt protein-coding genes. Numbers above branches represent support for MP (BPs; upper value of each quartet), ME (mean character distances; BPs; middle-upper value), ML (mtREV24 + Γ + I model; QPs; middle-lower value), and BI (mtREV24 + Γ + I model; BPPs; lower value). Hyphens indicate support values of less than 50. Numbers below branches represent decay indices. Lobe-finned fishes (*Protopterus* and *Latimeria*) were used as outgroups.

All phylogenetic analyses, MP (6229 steps; CI = 0.71), ME (score = 1.43), ML (−Ln likelihood = 38653.67), and BI (−Ln likelihood = 38676.48) yielded the same inferred relationships with differences only in branch lengths and levels of support (Fig. 5). With all methods and measures, quantitative support for the monophyly of living

amphibians (Lissamphibia), salamanders (Caudata), and caecilians is maximal, whereas for the monophyly of frogs (Anura) and the sister group relationship between Anura and Caudata (the Batrachia hypothesis; Milner, 1988) BPPs are maximal and QPs are high (>90), but BPs are less impressive (63–71). Within caecilians, quantitative support for the sister group pairing of *Ichthyophis* and *Uraeotyphlus* is maximal with all methods. *Rhinatrema* was recovered as the sister group of all other caecilians, with only ME providing support values less than 95. Only BPPs and QPs indicate strong support for a higher-caecilian clade, comprising *Typhlonectes*, *Gegeneophis*, and *Scolecomorphus*, and for a sister group pairing of *Typhlonectes* and *Gegeneophis*.

RAG1 gene nucleotide sequences produced a raw alignment of 1512 positions. One gapped codon was excluded, and of the remaining positions, 795 are invariant and 503 parsimony-informative. Uncorrected "*p*" distances among caecilian taxa ranged from 0.04 (*Ichthyophis* vs. *Uraeotyphlus*) to 0.11 (*Rhinatrema* vs. *Scolecomorphus*), and among amphibian orders from 0.22 to 0.25. Interestingly, only 90 RAG1 positions are parsimony-informative among the sampled caecilians. Using Modeltest, we selected the parameter-rich GTR (Rodríguez et al., 1990) + Γ + I model of substitution for the ML and ME analyses. For the Bayesian analyses, the best-fit models selected were GTR + Γ for the first and the second position partitions and GTR + Γ + I for third positions.

All methods, ML (−Ln likelihood = 8379.63), MP (1507 steps; CI = 0.69), BI (−Ln likelihood = 8071.51), and ME (score = 1.50), produced single trees that differed only in the interrelationships among the caecilian lineages (Fig. 6) and which are otherwise congruent with the single tree inferred from the mt data. As with the mt data, there is maximal or very strong support for the monophyly of Lissamphibia, Caudata, and Gymnophiona, but quantitative support for the Batrachia hypothesis is less impressive and appears strong only with BPPs and MP BPs (Fig. 6A). Within caecilians, only the sister group relationship of *Ichthyophis* and *Uraeotyphlus* was consistently recovered by all methods of analysis. As with the mt data, quantitative support for this relationship is maximal or nearly so. MP, ML, and BI on the one hand, and ME on the other, yielded two different trees for caecilians (Fig. 6B). Of these, the ME tree is most similar to that inferred from the mt data, differing from it only in the resolution of the relationships of the higher caecilians *Gegeneophis*, *Scolecomorphus*, and *Typhlonectes*. MP, ML, and BI yielded a tree (Fig. 6B) that conflicts dramatically with the mt data (Fig. 5). In all analyses, quantitative support values for the non-con-

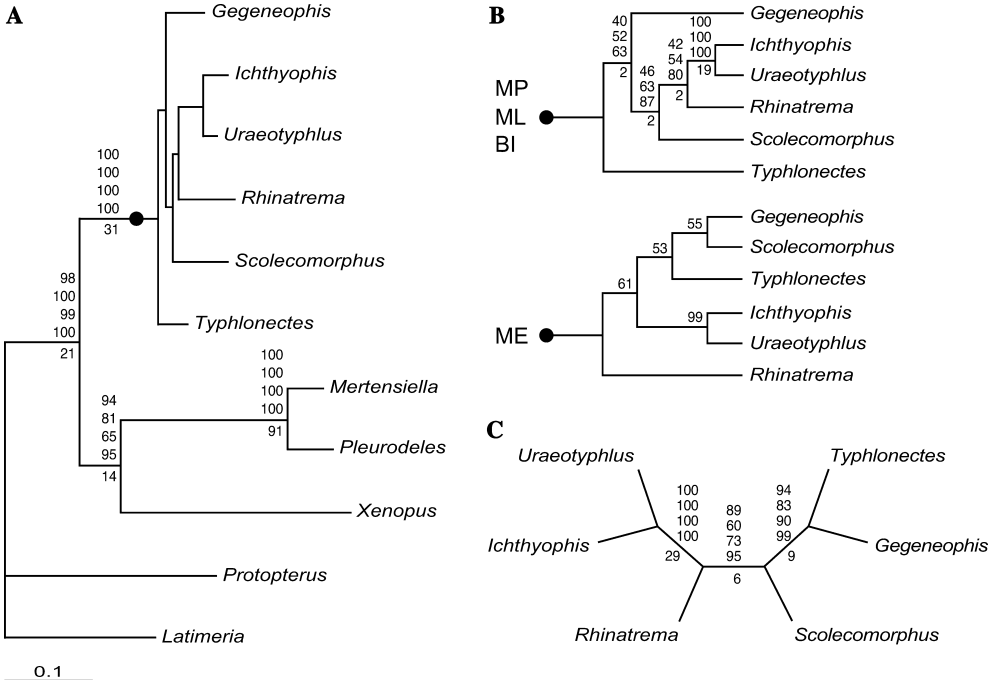


Fig. 6. Phylogenetic relationships of caecilian lineages inferred from RAG1 nucleotide sequence data. (A) ML phylogram showing statistical support for the clades outside Gymnophiona. Numbers above branches indicate support for MP (BPs; upper value of each quartet), ME (GTR + Γ + I distances; BPs; middle-upper value), ML (GTR + Γ + I model; BPs; middle-lower value), and BI (different models according to codon position, see text; BPPs; lower value). Lobe-finned fishes (*Protopterus* and *Latimeria*) were used as outgroups. (B) Inferred relationships and statistical support for each phylogenetic method for the clades within Gymnophiona. (C) Unrooted tree inferred from analysis of caecilian-only RAG1 data, with statistical support for the four methods given in same order as listed for (A). Numbers below branches represent decay indices in all trees.

gruent relationships are unimpressive. Much of the difference between the alternative trees for the RAG1 data is attributable to different rootings of the caecilian clade. To further explore this, we performed an unrooted analysis using the caecilian-only data. All methods yielded an unrooted tree (Fig. 6C) that is fully consistent with the tree supported by the mt data (Fig. 5) and measures of support are considerably enhanced compared to the corresponding splits in the analyses of the full RAG1 data.

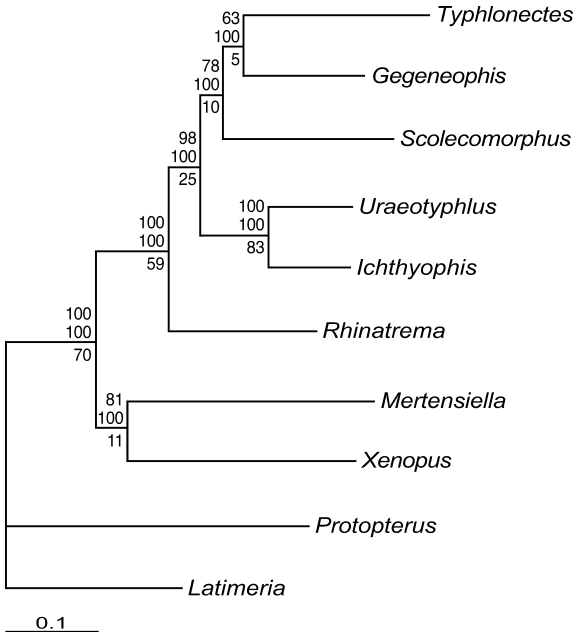


Fig. 7. Phylogenetic relationships (BI phylogram) of caecilians inferred from a combined data set of mt protein amino acids and RAG1 nucleotide sequences. Numbers above branches represent support for MP (BPs; upper value) and BI (mtREV24 + Γ + I model for the mt proteins, and different models according to codon position for the RAG1 gene, see text; BPPs; lower value). Numbers below branches represent decay indices. Lobe-finned fishes (*Protopterus* and *Latimeria*) were used as outgroups.

The combined mt and nuclear data comprised a total of 4991 sites. For BI, the best-fitting model for the mt amino acid partition was mtREV24 + Γ + I, and for the nucleotide RAG1 data were GTR + Γ for the first position, GTR + Γ + I for the second position, and GTR + Γ for the third position. BI and MP yielded the same tree (Fig. 7; −Ln likelihood = 41,461.80; 6551 parsimony steps, CI = 0.74). Relationships among the caecilians are identical to those recovered from the mt data using all methods (Fig. 5) and are fully consistent with the unrooted analysis of the RAG1 caecilian-only data. BPPs are maximal for all relationships in this tree. Parsimony BPs are maximal or nearly so for Lissamphibia, Gymnophiona, the sister group relationship of *Rhinatrema* to all other caecilians and the *Ichthyophis* + *Uraeotyphlus* pairing, substantial (>75) for the Batrachia hypothesis and the "higher" caecilians, but less impressive for the resolution of relationships within the higher caecilians.

Results of AU, SH, KH, and Templeton tests of alternative tree topologies, using caecilian-only data, are summarized in Table 3. Although unrooted, for convenience we describe these trees as if they were rooted on *Rhinatrema*. With either the mt or the RAG1 data, SH tests allow us to reject only trees that do not include the grouping of *Ichthyophis* and *Uraeotyphlus*. AU, KH, and Templeton tests also allow rejection of these trees but, in addition, allow rejection of some of the 15 trees

Table 3
Log likelihood and *p* values of Approximately Unbiased (AU), Shimodaira–Hasegawa (SH), Kishino–Hasegawa (KH), and Templeton tests for each of the 15 unrooted topologies that maintain the *Ichthyophis* + *Uraeotyphlus* pairing for the caecilian-only data

Alternative topologies	mt proteins					RAG1				
	−log L	AU	SH	KH	Templeton	−log L	AU	SH	KH	Templeton
(G,((((I,U),R),S),T))	23748.22	0.900	0.998	0.798	0.697	4094.83	0.902	0.994	0.781	1.000
(G,((((I,U),R),T),S))	23756.37	0.329	0.863	0.202	0.558	4101.83	0.093	0.677	0.077	0.029
(G,(((I,U),R),(S,T)))	23763.21	0.036	0.739	0.044	0.939	4105.61	0.094	0.677	0.077	0.012
(G,(((I,U),(R,S)),T))	23770.63	0.047	0.603	0.032	0.821	4106.19	0.080	0.810	0.126	0.033
(G,((((I,U),S),R),T))	23777.95	<0.001	0.471	0.003	0.125	4101.83	0.360	0.844	0.219	0.134
(G,((((I,U),T),R),S))	23778.06	0.041	0.467	0.040	0.287	4097.87	0.088	0.550	0.068	0.011
(G,(((I,U),(R,T)),S))	23779.09	0.036	0.456	0.033	0.346	4105.77	0.037	0.523	0.052	0.086
(G,(((I,U),T),(R,S)))	23786.58	0.032	0.335	0.015	0.286	4105.61	0.089	0.550	0.068	0.007
(G,((I,U),((R,S),T)))	23790.74	0.007	0.281	0.007	1.000	4107.66	0.004	0.493	0.031	0.007
(G,((I,U),((R,T),S)))	23791.23	0.024	0.275	0.007	0.431	4104.66	0.037	0.522	0.053	0.046
(G,((((I,U),T),S),R))	23793.72	0.002	0.247	0.004	0.158	4098.67	0.087	0.550	0.068	0.016
(G,((I,U),(R,(S,T))))	23794.61	0.002	0.236	0.003	0.305	4105.61	0.004	0.493	0.031	0.007
(G,(((I,U),S),(R,T)))	23796.24	0.002	0.216	0.002	0.127	4107.46	0.144	0.574	0.085	0.071
(G,(((I,U),(S,T)),R))	23799.82	0.005	0.177	0.001	0.031	4106.20	0.003	0.491	0.028	0.004
(G,((((I,U),S),T),R))	23802.56	0.007	0.156	<0.001	0.025	4107.46	0.034	0.541	0.057	0.009

The first topology corresponds to the optimal ML tree. Ninety of the 105 possible topologies are not shown because all four tests reject them at $p < 0.001$. G, *Gegeneophis*; I, *Ichthyophis*; R, *Rhinatrema*; S, *Scolecomorphus*; T, *Typhlonectes*; and U, *Uraeotyphlus*.

that place *Ichthyophis* and *Uraeotyphlus* together. With RAG1, AU tests allow us to reject six of these 15 trees, including all those in which *Gegeneophis* is more closely related to *Ichthyophis* and *Uraeotyphlus* than to any other caecilians. With the mt data, the AU tests allow us to reject all except one suboptimal tree, in which *Gegeneophis* is most closely related to *Scolecomorphus* rather than to *Typhlonectes*. KH test results match closely those of the AU test. Templeton test results match closely those of SH test with the mt data, but allow us to reject 11 of the 15 trees with RAG1.

4. Discussion

4.1. Distinct features of the new caecilian mitochondrial genomes

The new caecilian mt genomes are similar in size and gene arrangement to those of *T. natans* (Zardoya and Meyer, 2000), and thus conform to the vertebrate consensus organization (Jameson et al., 2003). The only exception is the mt genome of *G. ramaswamii*, which lacks the $tRNA^{Phe}$ gene. This presumably derived absence is unique among known vertebrate mt genomes. Absence of other tRNA genes has been previously reported in marsupials (Janke et al., 1997, 2002), and the tuatara (Rest et al., 2003). In marsupials, it has been shown that an alternative tRNA of nuclear origin is imported into mitochondria to participate in the translation process (Dorner et al., 2001). Given that the usage of phenylalanine in the mt proteins of *G. ramaswamii* is comparable to that in the other caecilians (not shown), an analogous importation may be implicated.

All caecilian mt control regions lack tandem repeats with the exception of that of *T. natans* (Zardoya and Meyer, 2000). The newly reported caecilian CSB-1 motifs are not reduced to a truncated pentamotif (5′-GACAT-3′) as in fishes (e.g., Hurst et al., 1999), but share high similarity with the mouse CSB-1 (Walberg and Clayton, 1981). A truncated CSB-1 was tentatively reported for *T. natans* (Zardoya and Meyer, 2000), but the alignment of all caecilian mt control regions allowed us to identify a complete CSB-1 motif in this species. One of the pyrimidine-rich regions, PP-1 (poly(T) stretch), has been previously described for several fishes (Hurst et al., 1999) and might be involved in regulatory aspects of the origin of H-strand replication. A second pyrimidine-rich region, PP-2 (poly(C) stretch), shows a moderately high similarity in most caecilians to the downstream CSB-2 motif, and could be the result of a former duplication. Except for *S. vittatus* and *T. natans*, the general absence of TAS (Doda et al., 1981) at the 5′ end of caecilian mt control regions contrasts with their presence and putative essential role in arresting replication in many other vertebrate mt genomes.

Unusually long intergenic spacers were found between the $tRNA^{Thr}$ and $tRNA^{Pro}$ genes in *R. bivittatum* and *U.* cf. *oxyurus*. Other cases of long intervening noncoding sequences have been reported in salamanders (Zardoya et al., 2003; Zhang et al., 2003), but not in frogs to date, and in all cases sequence similarities are low, suggesting that they are not homologous.

4.2. Comparative analysis of mitochondrial and RAG1 molecular features

Overall base compositions are biased against guanine in all caecilian mt genomes (not shown). This is a typical feature of vertebrate mtDNA, and is mainly due to a strong selection against the use of guanine at third codon positions of protein-coding genes (Zardoya and Meyer, 2000). In contrast, RAG1 gene sequences are homogenous in base composition, similar to what has been reported for this gene in birds and crocodiles (Groth and Barrowclough, 1999). The only exceptions are the salamander RAG1 sequences, which show a moderately strong G+C bias. This may reflect constraints in codon usage in this amphibian group. More amphibian RAG1 gene sequences (especially from anurans and salamanders) need to be determined to further investigate this distinctive condition.

Substitution rate of RAG1 at both nucleotide and amino acid levels was relatively slower than that of almost all mt proteins, being similar to those of the amino acid sequences of the most conservative mt protein-coding genes (cytochrome oxidase subunits, Zardoya and Meyer, 1996b). This makes RAG1 a potentially useful molecular marker for the study of deep vertebrate divergences. Among-site rate heterogeneity of RAG1 is quite similar to those of most mt proteins, being a little higher at the nucleotide level. Only ATP8 shows an unexpectedly high value of among-site rate heterogeneity, which is consistent with the fact that this is the shortest mt protein-coding gene, and shows few conserved positions across vertebrates (Zardoya and Meyer, 1996b).

4.3. Phylogenetics

All relevant analyses provide strong support for four uncontroversial high-level relationships—monophyly of Lissamphibia, Anura, Caudata, and Gymnophiona. Additionally, all relevant analyses are consistent with the Batrachia hypothesis (Gymnophiona, (Anura, Caudata)). This resolution of the Lissamphibia problem is the best hypothesis given the available data, but support is not consistently high in all analyses and, given the limited sampling of anuran and caudate taxa, it cannot be accepted without reservation.

With the exception of RAG1 only, all analyses strongly support the conventional view based on morphology that the Rhinatrematidae is the sister group

of all other caecilians. With outgroups included, RAG1 supports two alternative caecilian trees depending on the method of analysis. Only ME produces the expected sister group relationship between *Rhinatrema* and the other caecilians, with other methods recovering *Typhlonectes* as the sister taxon of other caecilians. In neither arrangement are support values for the basal split high, and low support also characterizes the other intracaecilian relationships except for the pairing of *Uraeotyphlus* and *Ichthyophis*. The latter is extremely well supported in all analyses, in statistical tests, and by previous analyses of morphology and molecules (see below). When outgroups are excluded from RAG1 analyses, all methods yield a single unrooted caecilian tree, fully consistent with relationships inferred from mt amino acid sequence data. This suggests that when outgroups are included there is a local rooting problem in the caecilian tree. Wilkinson (1996b) showed how unstable leaves (taxa) can decrease the bootstrap support for otherwise well-supported relationships. The present example is a special case in which the unstable "leaf" is the root of the caecilian tree. In order to investigate whether low support for the Batrachia hypothesis was a product of the instability of the root of the caecilian tree, we repeated the RAG1 analyses after exclusion of the higher caecilians. This had no substantial impact upon support values (not shown).

The sister group relationship between *Uraeotyphlus* and *Ichthyophis* appears to be the best supported relationship among the sampled caecilians. It is recovered in all analyses with maximal or near maximal support. Monophyly of the higher caecilians (*Gegeneophis*, *Scolecomorphus*, and *Typhlonectes*) is also supported in all analyses that were not affected by local rooting problems but measures of support, though generally high, are not universally high. Most uncertainty remains in the resolution of the higher caecilians. Mt amino acid and combined analyses place *Gegeneophis* with *Typhlonectes*, but with a mixture of high (BPP and QP) and low (MP and ME BPs) support.

There appears to be considerable uncertainty as to which of the various parametric and non-parametric likelihood-based tests are best used to determine whether the difference in fit of two or more trees to the data is significantly greater than expected under the null hypothesis of random sampling error. Although the KH test has been widely used, its validity requires the trees to be specified a priori rather than chosen on the basis of their likelihoods (Goldman et al., 2000). The SH test can be used to evaluate trees chosen a posteriori, but to be valid it requires the inclusion of all "reasonable" trees, and it is unclear how the set of reasonable trees can be selected (Buckley, 2002). On this point, Goldman et al. (2000) note only that selecting all possible trees will always be conservative, but this is an impractical selection for all but the smallest taxon samplings. Empirical comparisons of non-parametric SH tests, and of tests that use parametric bootstrapping, have provided very divergent results and quite different biological conclusions that suggest the SH test is very conservative because of the multiple comparisons, and that parametric bootstrapping may be too liberal as a result of model misspecification (e.g., Buckley, 2002; Goldman et al., 2000; Strimmer and Rambaut, 2001). A further uncertainty arises when the trees to be compared are chosen partly a priori and partly a posteriori, such as when we are interested in a putative monophyletic group but not in the resolution of relationships within that group. The more recently developed AU test is non-parametric and uses a multiscale bootstrap approach. It is less biased than other methods, but is also impractical when the number of trees to be compared is large (Shimodaira, 2002).

We used multiple non-parametric likelihood-based KH, SH, and AU tests and parsimony-based Templeton tests to further evaluate the strength of our inferences on caecilian relationships, and to provide an empirical comparison of the tests. We used the caecilian-only data because for six taxa there are only 105 possible unrooted trees, making the selection of all possible trees practical. Based on previous analyses of morphology and mt DNA sequence data, we expect two splits to be present in the caecilian tree, the pairing of *Uraeotyphlus* and *Ichthyophis*, and the partitioning of the higher caecilians, with particularly strong prior confidence in the former. Thus, based on a priori considerations, we are interested in comparing the three alternative resolutions of the higher caecilians. With RAG1, KH tests do not allow us to reject any of these alternatives, whereas with the more substantial mt amino acid data, KH tests allow the rejection of the grouping of *Scolecomorphus* with *Typhlonectes*, leaving a pairing of *Gegeneophis* with either *Typhlonectes*, as in the optimal tree, or with *Scolecomorphus* as viable alternative hypotheses. Ignoring a priori expectations and examining all 105 possible trees, SH tests are much less discriminatory. Using RAG1 or mt amino acid data, SH tests agree in rejecting only those trees that do not include the pairing of *Uraeotyphlus* and *Ichthyophis*. Using AU tests, these trees are also rejected, but RAG1 allows rejection of six additional trees (those that place *Gegeneophis* in a partition with *Uraeotyphlus* and *Ichthyophis*) and mt amino acid data allow rejection of all except one suboptimal tree (that placing *Gegeneophis* with *Scolecomorphus*).

Although the a posteriori SH tests provide strong support for our a priori confidence in the *Uraeotyphlus* and *Ichthyophis* pairing, the failure to discriminate against other hypotheses is disappointing given the amount of data available and the levels of support indicated by BPs, BPPs, and QPs. This suggests that the conservative SH test is too conservative.

Comparative results from the AU test bear this out. AU tests indicate that the mt amino acid data, in particular, strongly support the higher caecilian grouping also, and fail only to discriminate between the placement of *Gegeneophis* with *Typhlonectes* or with *Scolecomorphus* within the higher caecilians. Despite the concern that KH tests of trees that are not selected a priori are biased (Goldman et al., 2000), our KH test results are very similar to those obtained using the AU test. Good, but less tight correlation between KH and AU test results are reported for other data sets by Shimodaira (2002). The extent to which easily implemented KH tests may be a reasonable proxy for the more computationally demanding AU tests merits further investigation. Templeton test results are quite dissimilar between mt data and RAG1, seeming too conservative with the former and highly discriminative with the latter.

To summarize our phylogenetic investigations (using the initial letters to represent genera), we consider that the mt and nuclear data provide good support for (R,((I,U),(S,(G,T)))), and this is our preferred tree, although we do not discount (R,((I,U),(T,(S,G)))). Assuming that *Caecilia*, the type genus of the Caeciliidae, is more closely related to *Typhlonectes* than is *Gegeneophis* (Wilkinson et al., 2003) the latter tree would indicate that the Caeciliidae is paraphyletic with respect to the Scolecomorphidae as well as with respect to the Typhlonectidae.

Using mt ribosomal DNA sequence data, Wilkinson et al. (2003) were unable to resolve a number of relationships among the sampled caecilians. Our analyses demonstrate the potential of both mt protein gene and nuclear RAG1 data for providing well-supported resolution of caecilian phylogenetic relationships. Thus, expanded taxon sampling for these data is expected to provide much needed additional insights into caecilian phylogeny, particularly with respect to poorly understood relationships among the higher caecilians.

Acknowledgments

Lukas Rüber provided helpful technical advice with laboratory work and phylogenetic analyses. Peter Foster kindly assisted with the AU, KH, and SH tests. In addition to the many people acknowledged by Gower et al. (2002), Wilkinson et al. (2002, 2003), and Gower et al. (2004) for their help with fieldwork, we thank Tito Barros and Simon Loader. Two anonymous reviewers gave insightful comments on an earlier version of the manuscript. D.S.M. was sponsored by a predoctoral fellowship of the Ministerio de Ciencia y Tecnología of Spain. D.J.G and M.W. were granted a BIODIBERIA award to visit the MNCN. This work received financial support from a project of the Ministerio de Ciencia y Tecnología of Spain to R.Z. (CGL2004-00401). Fieldwork was funded in part by NERC GST/02/832.

References

Adachi, J., Hasegawa, M., 1996. Model of amino acid substitution in proteins encoded by mitochondrial DNA. J. Mol. Evol. 42, 459–468.

Brinkmann, H., Venkatesh, B., Brenner, S., Meyer, A., 2004. Nuclear protein-coding genes support lungfish and not the coelacanth as the closest living relatives of land vertebrates. Proc. Natl. Acad. Sci. USA 101, 4900–4905.

Buckley, T.R., 2002. Model misspecification and probabilistic tests of topology: evidence from empirical data sets. Syst. Biol. 51, 509–523.

Castresana, J., 2000. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. Mol. Biol. Evol. 17, 540–552.

Cummings, M.P., Otto, S.P., Wakeley, J., 1995. Sampling properties of DNA sequence data in phylogenetic analysis. Mol. Biol. Evol. 12, 814–822.

Curole, J.P., Kocher, T.D., 1999. Mitogenomics: digging deeper with complete mitochondrial genomes. Trends Ecol. Evol. 14, 394–398.

Doda, J.N., Wright, C.T., Clayton, D.A., 1981. Elongation of displacement loop strands in human and mouse mitochondrial DNA is arrested near specific template sequences. Proc. Natl. Acad. Sci. USA 78, 6116–6120.

Dorner, M., Altmann, M., Pääbo, S., Morl, M., 2001. Evidence for import of a lysyl-tRNA into marsupial mitochondria. Mol. Biol. Cell 12, 2688–2698.

Duellman, W.E., Trueb, L., 1986. Biology of Amphibians. McGraw-Hill, New York.

Eriksson, T., 2001. AutoDecay. Available from: <http://www.bergianska.se/index_forskning_soft.html>.

Evans, S.E., Sigogneau-Russel, D., 2001. A stem-group caecilian (Lissamphibia: Gymnophiona) from the Lower Cretaceous of North Africa. Palaeontology 44, 259–273.

Felsenstein, J., 1981. Evolutionary trees from DNA sequences: a maximum likelihood approach. J. Mol. Evol. 17, 368–376.

Fitch, W.M., 1971. Toward defining the course of evolution: minimal change for a specific tree topology. Syst. Zool. 20, 406–416.

Foster, P., 2003. p4. Available form: <http://www.nhm.ac.uk/zoology/external/p4.htm>.

Frippiat, C., Kremarik, P., Ropars, A., Dournon, C., Frippiat, J.P., 2001. The recombination-activating gene 1 of *Pleurodeles waltl* (urodele amphibian) is transcribed in lymphoid tissues and in the central nervous system. Immunogenetics 52, 264–275.

Frost, D.R., 2002. Amphibian species of the World: an online reference. V2.21 (15 July, 2002). Available from: <http://research.amnh.org/herpetology/amphibia/index.html>.

Goldman, N., Anderson, J.P., Rodrigo, A.G., 2000. Likelihood-based tests of topologies in phylogenetics. Syst. Biol. 49, 652–670.

Gower, D.J., Kupfer, A., Oommen, O.V., Himstedt, W., Nussbaum, R.A., Loader, S.P., Presswell, B., Müller, H., Krishna, S.B., Boistel, R., Wilkinson, M., 2002. A molecular phylogeny of ichthyophiid caecilians (Amphibia: Gymnophiona: Ichthyophiidae): out of India or out of South East Asia?. Proc. R. Soc. Lond. B 269, 1563–1569.

Gower, D.J., Loader, S.P., Moncrieff, C.B., Wilkinson, M., 2004. Niche separation and comparative abundance of *Boulengerula boulengeri* and *Scolecomorphus vittatus* (Amphibia: Gymnophiona) in an East Usambara forest, Tanzania. Afr. J. Herpetol. (in press).

Greenhalgh, P., Olesen, C.E., Steiner, L.A., 1993. Characterization and expression of recombination activating genes (RAG-1 and RAG-2) in *Xenopus laevis*. J. Immunol. 151, 3100–3110.

Groth, J.G., Barrowclough, G.F., 1999. Basal divergences in birds and the phylogenetic utility of the nuclear RAG-1 gene. Mol. Phylogenet. Evol. 12, 115–123.

Hay, J.M., Ruvinsky, I., Hedges, S.B., Maxson, L.R., 1995. Phylogenetic relationships of amphibian families inferred from DNA sequences of mitochondrial 12S and 16S ribosomal RNA genes. Mol. Biol. Evol. 12, 928–937.

Hedges, S.B., Maxson, L.R., 1993. A molecular perspective on lissamphibian phylogeny. Herpetol. Monogr. 7, 27–42.

Hedges, S.B., Nussbaum, R.A., Maxson, L.R., 1993. Caecilian phylogeny and biogeography inferred from mitochondrial DNA sequences of the 12SrRNA and 16S rRNA genes (Amphibia: Gymnophiona). Herpetol. Monogr. 7, 64–76.

Hillis, D.M., 1991. The phylogeny of amphibians: current knowledge and the role of cytogenetics. In: Sessions, S.K., Green, D.M. (Eds.), Amphibian Cytogenetics and Evolution. Academic Press, San Diego, CA, pp. 7–31.

Himstedt, W., 1996. Die Blindwühlen. Westarp Wissenschaften, Magdeburg.

Hixson, J.E., Wong, T.W., Clayton, D.A., 1986. Both the conserved stem-loop and divergent 5′-flanking sequences are required for initiation at the human mitochondrial origin of light-strand DNA replication. J. Biol. Chem. 261, 2384–2390.

Huelsenbeck, J.P., Ronquist, F.R., 2001. MrBayes: Bayesian inference of phylogeny. Bioinformatics 17, 754–755.

Huelsenbeck, J.P., Ronquist, F.R., Nielsen, R., Bollback, J.P., 2001. Bayesian inference of phylogeny and its impact on evolutionary biology. Science 294, 2310–2314.

Hurst, C.D., Bartlett, S.E., Davidson, W.S., Bruce, I.J., 1999. The complete mitochondrial DNA sequence of the Atlantic salmon, *Salmo salar*. Gene 239, 237–242.

Jameson, D., Gibson, A.P., Hudelot, C., Higgs, P.G., 2003. OGRe: a relational database for comparative analyses of mitochondrial genomes. Nucleic Acids Res. 31, 202–206.

Janke, A., Magnell, O., Wieczorek, G., Westerman, M., Arnason, U., 2002. Phylogenetic analyses of 18S rRNA and the mitochondrial genomes of the Wombat, *Vombatus ursinus*, and the spiny anteater, *Tachyglossus aculeatus*: increased support for the Marsupionta hypothesis. J. Mol. Evol. 54, 71–80.

Janke, A., Xu, X., Arnason, U., 1997. The complete mitochondrial genome of the wallaroo (*Macropus robustus*) and the phylogenetic relationships among Monotremata, Marsupialia, and Eutheria. Proc. Natl. Acad. Sci. USA 94, 1276–1281.

Kishino, H., Hasegawa, M., 1989. Evaluation of the maximum likelihood estimate of the evolutionary tree topologies from DNA sequence data, and the branching order in Hominoidea. J. Mol. Evol. 29, 170–179.

Kocher, T.D., Thomas, W.K., Meyer, A., Edwards, S.V., Pääbo, S., Villablanca, F.X., Wilson, A.C., 1989. Dynamics of mitochondrial DNA evolution in animals: amplification and sequencing with conserved primers. Proc. Natl. Acad. Sci. USA 86, 6196–6200.

Loader, S.P., Wilkinson, M., Gower, D.J., Msuya, C.A., 2003. A remarkable young *Scolecomorphus vittatus* (Amphibia: Gymnophiona: Scolecomorphidae) from the North Pare Mountains, Tanzania. J. Zool. 259, 93–101.

Martin, A.P., 1999. Substitution rates of organelle and nuclear genes in sharks: implicating metabolic rate (again). Mol. Biol. Evol. 16, 996–1002.

Measey, G.J., Gower, D.J., Oommen, O.V., Wilkinson, M., 2003. Quantitative surveying of limbless endogenic vertebrates—a case study of *Gegeneophis ramaswamii* (Amphibia: Gymnophiona) in southern India. Appl. Soil Ecol. 23, 43–53.

Milner, A.R., 1988. The relationships and origin of living amphibians. In: Benton, M.J. (Ed.), The Phylogeny and Classification of the Tetrapods. Clarendon Press, Oxford, pp. 59–102.

Noble, G.K., 1931. The Biology of the Amphibia. McGraw-Hill, New York.

Nussbaum, R.A., 1977. Rhinatrematidae: a new family of caecilians (Amphibia: Gymnophiona). Occ. Pap. Mus. Zool. Univ. Michigan 682, 1–30.

Nussbaum, R.A., 1979. The taxonomic status of the caecilian genus *Uraeotyphlus* Peters. Occ. Pap. Mus. Zool. Univ. Michigan 687, 1–20.

Nussbaum, R.A., 1985. Systematics of the caecilians (Amphibia: Gymnophiona) of the family Scolecomorphidae. Occ. Pap. Mus. Zool. Univ. Michigan 713, 1–49.

Nussbaum, R.A., 1991. Cytotaxonomy of caecilians. In: Sessions, S.K., Green, D.M. (Eds.), Amphibian Cytogenetics and Evolution. Academic Press, San Diego, CA, pp. 22–76.

Nussbaum, R.A., Wilkinson, M., 1989. On the classification and phylogeny of caecilians (Amphibia: Gymnophiona), a critical review. Herpetol. Monogr. 3, 1–42.

Palumbi, S.R., Martin, A., Romano, S., Owen MacMillan, W., Stice, L., Grabowski, G., 1991. The Simple Fool's Guide to PCR. Department of Zoology, University of Hawaii, Honolulu.

Posada, D., Crandall, K.A., 1998. Modeltest: testing the model of DNA substitution. Bioinformatics 14, 817–818.

Rest, J.S., Ast, J.C., Austin, C.C., Waddell, P.J., Tibbetts, E.A., Hay, J.M., Mindell, D.P., 2003. Molecular systematics of primary reptilian lineages and the tuatara mitochondrial genome. Mol. Phylogenet. Evol. 29, 289–297.

Rodríguez, F., Oliver, J.F., Marín, A., Medina, J.R., 1990. The general stochastic model of nucleotide substitution. J. Theor. Biol. 142, 485–501.

Roe, B.A., Din-Pow, M., Wilson, R.K., Wong, J.F., 1985. The complete nucleotide sequence of the *Xenopus laevis* mitochondrial genome. J. Biol. Chem. 260, 9759–9774.

Russo, C.A.M., Takezaki, N., Nei, M., 1996. Efficiencies of different genes and different tree-building methods in recovering a known vertebrate phylogeny. Mol. Biol. Evol. 13, 525–536.

Rzhetsky, A., Nei, M., 1992. A simple method for estimating and testing minimum-evolution trees. Mol. Biol. Evol. 9, 945–967.

Sambrook, J., Fritsch, E.F., Maniatis, T., 1989. Molecular Cloning. A Laboratory Manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Shimodaira, H., 2002. An approximately unbiased test of phylogenetic tree selection. Syst. Biol. 51, 492–508.

Shimodaira, H., Hasegawa, M., 1999. Multiple comparisons of Log-likelihoods with applications to phylogenetic inference. Mol. Biol. Evol. 16, 1114–1116.

Shimodaira, H., Hasegawa, M., 2001. Consel: for assessing the confidence of phylogenetic tree selection. Bioinformatics 17, 1246–1247.

StatSoft Inc., 2001. STATISTICA (data analysis software system). http://www.statsoft.com.

Strimmer, K., Rambaut, A., 2001. Inferring confidence sets of possible misspecified gene trees. Proc. R. Soc. Lond. B 269, 137–142.

Strimmer, K., von Haeseler, A., 1996. Quartet puzzling: a quartet maximum-likelihood method for reconstructing tree topologies. Mol. Biol. Evol. 13, 964–969.

Sumida, M., Kanamori, Y., Kaneda, H., kato, Y., Nishioka, M., Hasegawa, M., Yonekawa, H., 2001. Complete nucleotide sequence and gene rearrangement of the mitochondrial genome of the Japanese pond frog *Rana nigromaculata*. Genes Genet. Syst. 76, 311–325.

Swofford, D.L., 1998. PAUP*: phylogenetic analysis using parsimony (*and other methods), version 4.0. Sinauer Associates. Inc, Sunderland, MA, USA.

Taylor, E.H., 1968. The Caecilians of the World: A Taxonomic Analysis. University of Kansas Press, Lawrence, KS.

Taylor, E.H., 1969. A new family of African Gymnophiona. Univ. Kansas Sci. Bull. 48, 297–305.

Templeton, A.R., 1983. Phylogenetic inference from restriction endonuclease cleavage site maps with particular reference to the evolution of human and the apes. Evolution 37, 221–244.

Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmougin, J., Higgins, D.G., 1997. The CLUSTAL X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res. 25, 4876–4882.

Wake, M.H., 1977. The reproductive biology of caecilians: an evolutionary perspective. In: Taylor, E.H., Guttman, S.I. (Eds.), Reproductive Biology of Amphibians. Plenum Press, New York, pp. 73–101.

Walberg, M.W., Clayton, D.A., 1981. Sequence and properties of the human KB cell and mouse L cell D-loop regions of mitochondrial DNA. Nucleic Acids Res. 9, 5411–5421.

Wilkinson, M., 1992. The phylogenetic position of the Rhinatrematidae (Amphibia: Gymnophiona): evidence from the larval lateral line system. Amphibia–Reptilia 13, 74–79.

Wilkinson, M., 1996a. The heart and aortic arches of rhinatrematid caecilians (Amphibia: Gymnophiona). Zoomorphology 105, 277–295.

Wilkinson, M., 1996b. Majority-rule reduced consensus methods and their use in bootstrapping. Mol. Biol. Evol. 13, 437–444.

Wilkinson, M., 1997. Characters, congruence and quality: a study of neuroanatomical and traditional data in caecilian phylogeny. Biol. Rev. 72, 423–470.

Wilkinson, M., Loader, S.P., Gower, D.J., Sheps, J.A., Cohen, B.L., 2003. Phylogenetic relationships of African caecilians (Amphibia: Gymnophiona): insights from mitochondrial rRNA gene sequences. Afr. J. Herpetol. 52, 83–92.

Wilkinson, M., Nussbaum, R.A., 1996. On the phylogenetic position of the Uraeotyphlidae (Amphibia: Gymnophiona). Copeia 1996, 550–562.

Wilkinson, M., Nussbaum, R.A., 1997. Comparative morphology and evolution of the lungless caecilian *Atretochoana eiselti* (Taylor) (Amphibia: Gymnophiona: Typhlonectidae). Biol. J. Linn. Soc. 62, 39–109.

Wilkinson, M., Nussbaum, R.A., 1999. Evolutionary relationships of the lungless caecilian *Atretochoana eiselti* (Amphibia: Gymnophiona: Typhlonectidae). Zool. J. Linn. Soc. 126, 191–223.

Wilkinson, M., Sheps, J.A., Oommen, O.V., Cohen, B.L., 2002. Phylogenetic relationships of Indian caecilians (Amphibia: Gymnophiona) inferred from mitochondrial rRNA gene sequences. Mol. Phylogenet. Evol. 23, 401–407.

Yang, Z., Nielsen, R., Hasegawa, M., 1998. Models of amino acid substitution and applications to mitochondrial protein evolution. Mol. Biol. Evol. 15, 1600–1611.

Zardoya, R., Meyer, A., 1996a. The complete nucleotide sequence of the mitochondrial genome of the lungfish (*Protopterus dolloi*) supports its phylogenetic position as a close relative of land vertebrates. Genetics 142, 1249–1263.

Zardoya, R., Meyer, A., 1996b. Phylogenetic performance of mitochondrial protein-coding genes in resolving relationships among vertebrates. Mol. Biol. Evol. 13, 933–942.

Zardoya, R., Meyer, A., 1997. The complete DNA sequence of the mitochondrial genome of a "living fossil," the coelacanth (*Latimeria chalumnae*). Genetics 146, 995–1010.

Zardoya, R., Meyer, A., 2000. Mitochondrial evidence on the phylogenetic position of Caecilians (Amphibia: Gymnophiona). Genetics 155, 765–775.

Zardoya, R., Meyer, A., 2001. On the origin of and phylogenetic relationships among living amphibians. Proc. Natl. Acad. Sci. USA 98, 7380–7383.

Zardoya, R., Malaga-Trillo, E., Veith, M., Meyer, A., 2003. Complete nucleotide sequence of the mitochondrial genome of a salamander, *Mertensiella luschani*. Gene 317, 17–27.

Zhang, P., Chen, Y.Q., Zhou, H., Wang, X.L., Qu, L.H., 2003. The complete mitochondrial genome of a relic salamander, *Ranodon sibiricus* (Amphibia: Caudata) and implications for amphibian phylogeny. Mol. Phylogenet. Evol. 28, 620–626.

5. PUBLICATION III

Title: **A hotspot of gene order rearrangement by tandem duplication and random loss in the vertebrate mitochondrial genome**

Authors: Diego San Mauro, David J. Gower, Rafael Zardoya, Mark Wilkinson

Status: Published

Year: 2006

Journal: Molecular Biology and Evolution (SCI Impact Factor: 6.355)

Volume: 23

Pages: 227-234

Resumen III (Spanish translation of the abstract of Publication III)

La mayoría de ejemplos de cambio en el orden de genes mitocondriales (mt) de vertebrados descritos pueden ser explicados por duplicación en tándem seguida de pérdida aleatoria de genes (modelo "*tandem duplication-random loss*" [TDRL]). Bajo este modelo de evolución, se predicen pérdidas independientes de genes procedentes de una única duplicación en una especie ancestral, y es esperable que puedan quedar pseudogenes (estadíos intermedios) en los genomas reordenados. Sin embargo, son raras las evidencias de esto, y están muy dispersas por los linajes de vertebrados. Aquí, describimos nuevos órdenes de genes mt derivados en la región "WANCY" de los vertebrados para cuatro cecilias estrechamente relacionadas. Los nuevos órdenes encontrados en esta región genómica (uno de ellos convergente con el órdenes derivado de marsupiales), la presencia de pseudogenes, y la posición de los espaciadores intergénicos satisfacen totalmente las predicciones del modelo TDRL. Nuestros resultados, junto con datos comparativos para otros genomas mt de vertebrados disponibles, proporcionan evidencia adicional de que la región genómica WANCY es un punto con alta tasa de reordenación de genes, y apoyan la idea de que el TDRL es el mecanismo dominante de reordenación de genes en el genoma mt de vertebrados. Las reordenaciones de genes convergentes no son improbables en puntos con alta tasa de reordenación de genes por TDRL.

A Hotspot of Gene Order Rearrangement by Tandem Duplication and Random Loss in the Vertebrate Mitochondrial Genome

Diego San Mauro, David J. Gower,† Rafael Zardoya,* and Mark Wilkinson†*

*Departamento de Biodiversidad y Biología Evolutiva, Museo Nacional de Ciencias Naturales, CSIC, Madrid, Spain; and †Department of Zoology, The Natural History Museum, London, United Kingdom

Most reported examples of change in vertebrate mitochondrial (mt) gene order could be explained by a tandem duplication followed by random loss of redundant genes (tandem duplication–random loss [TDRL] model). Under this model of evolution, independent loss of genes arising from a single duplication in an ancestral species are predicted, and remnant pseudogenes expected, intermediate states that may remain in rearranged genomes. However, evidence for this is rare and largely scattered across vertebrate lineages. Here, we report new derived mt gene orders in the vertebrate "WANCY" region of four closely related caecilian amphibians. The novel arrangements found in this genomic region (one of them is convergent with the derived arrangement of marsupials), presence of pseudogenes, and positions of intergenic spacers fully satisfy predictions from the TDRL model. Our results, together with comparative data for the available vertebrate complete mt genomes, provide further evidence that the WANCY genomic region is a hotspot for gene order rearrangements and support the view that TDRL is the dominant mechanism of gene order rearrangement in vertebrate mt genomes. Convergent gene rearrangements are not unlikely in hotspots of gene order rearrangement by TDRL.

Introduction

Most animal mitochondrial (mt) genomes studied contain the same 37 genes (Boore 1999; Jameson et al. 2003), but their order is variable among and, to a lesser extent, within major groups. Of the several mechanisms proposed to explain gene order rearrangements (e.g., Moritz and Brown 1986; Pääbo et al. 1991; Macey et al. 1997), tandem duplication followed by random gene loss is generally considered the most important in vertebrates (e.g., Moritz and Brown 1986, 1987; Moritz, Dowling, and Brown 1987; Pääbo et al. 1991; Arndt and Smith 1998; Boore 2000; Inoue et al. 2003). However, evidence for this in the form of duplicated genes that either remain functional or have become pseudogenes in the process of being eliminated is rather limited (Arndt and Smith 1998; Kumazawa et al. 1998; Macey et al. 1998; Liu, Wang, and Su 2005; Mueller and Boore 2005; Zhang et al. 2005), and most quantitative methods for the phylogenetic analysis of gene order data assume other rearrangement mechanisms (e.g., Sankoff et al. 1992; Blanchette, Kunisawa, and Sankoff 1999; Cosner et al. 2000; Larget, Kadane, and Simon 2005; Larget et al. 2005).

According to the tandem duplication–random loss (TDRL) model, novel gene orders result from random deletion of one of each of the pairs of the redundant paralogs produced by a tandem duplication (Moritz, Dowling, and Brown 1987; Boore 2000). Which gene is lost is determined by the accumulation of random (but see Lavrov, Boore, and Brown 2002) mutations that disrupt normal function and create a pseudogene that is further selected against and eventually lost from the genome. Alternative mechanisms including inversion (Smith et al. 1989), transposition (Macey et al. 1997), and intramolecular recombination (Lunt and Hyman 1997) have been suggested and sometimes invoked to account for mt gene order rearrangements that cannot be explained by TDRL alone (e.g., change in encoding strand requires some inversion), particularly in invertebrates (Dowton, Castro, and Austin 2002). Importantly, none of these alternative mechanisms explains the existence of pseudogenes, which require at least one duplication and that are expected intermediate steps in changing mt gene orders under TDRL (Macey et al. 1998).

We here report new data for the "WANCY" genomic region (including one new complete mt genome) of four closely related South American caecilian amphibians (Gymnophiona), three of the five nominate species of *Siphonops* and the closely related (Taylor 1968; Wilkinson and Nussbaum 1992) monotypic *Lutkenotyphlus*. These caecilians present novel arrangements of this region, presence of pseudogenes, and positions of intergenic spacers that fully satisfy predictions from the TDRL model. Our results and comparisons across the available mt gene order data for 453 vertebrates provide further evidence that the WANCY region is a hotspot for gene order rearrangements by TDRL (Boore and Brown 1998) and suggest that TDRL has been the principal mechanism of gene order rearrangement operating in the history of the vertebrate mt genome.

Materials and Methods

Taxon Sampling and DNA Sequencing

We determined the nucleotide sequence of the complete mt genome of the caecilian amphibian *Siphonops annulatus* and an mtDNA fragment that covered the WANCY region and part of flanking genes in two other species of *Siphonops* (*Siphonops paulensis* and *Siphonops hardyi*) and in *Lutkenotyphlus brasiliensis*. The WANCY region is a cluster of five tRNA genes ($tRNA^{Trp}$, $tRNA^{Ala}$, $tRNA^{Asn}$, $tRNA^{Cys}$, and $tRNA^{Tyr}$) surrounding the origin of light-strand replication (O_L) that is located between the genes for nicotinamide adenine dinucleotide dehydrogenase subunit 2 (*ND2*) and cytochrome c oxidase subunit 1 (*COX1*) in almost all vertebrate mt genomes (Seutin et al. 1994; Boore 1999; Jameson et al. 2003).

Caecilians (order Gymnophiona) are limbless, elongate amphibians distributed throughout mostly tropical habitats in Africa, America, and Asia (Taylor 1968; Duellman and Trueb 1994). All caecilian species examined

Key words: gene rearrangement, tandem duplication, gene loss, mitochondrial genome, convergent rearrangement, Gymnophiona.

E-mail: diegos@mncn.csic.es.

Mol. Biol. Evol. 23(1):227–234. 2006
doi:10.1093/molbev/msj025
Advance Access publication September 21, 2005

Table 1
Data for Caecilian Samples Employed in This Study

Species	Region Determined	Voucher Number	Collection Locality	GenBank Accession Number
Siphonops annulatus	Complete mt genome	BMNH 2005.9	Dominguez Martins, ES, Brazil	AY954506
Siphonops paulensis	WANCY region	CHUNB 39114	Formosa, GO, Brazil	AY954507
Siphonops hardyi	WANCY region	BMNH 2005.6	Dominguez Martins, ES, Brazil	AY954508
Lutkenotyphlus brasiliensis	WANCY region	BMNH 2005.3	Sao Paulo, SP, Brazil	AY954509

NOTE.—BMNH, The Natural History Museum, London; CHUNB, Departamento de Ciências Fisiológicas, Universidade de Brasilia.

in this study belong to the so-called "higher" caecilians (Nussbaum 1991), a well-defined clade that comprises three families (Caeciliidae, Scolecomorphidae, and Typhlonectidae) of still poorly known inter- and intrafamilial phylogenetic relationships (Wilkinson 1997; Wilkinson et al. 2003; San Mauro et al. 2004, 2005).

In all cases, total DNA was purified from ethanol-preserved liver or muscle with standard phenol/chloroform extraction procedures (Sambrook, Fritsch, and Maniatis 1989), and nucleotide sequences were determined using the primers, conditions, and methods reported by San Mauro et al. (2004). Details of the employed taxa, region sequenced, voucher specimens, collection localities, and GenBank accession numbers can be found in table 1.

The sequences of other available higher caecilians (Zardoya and Meyer 2000; San Mauro et al. 2004) were used as outgroups in phylogenetic analyses (GenBank accession numbers in parentheses): *Gegeneophis ramaswamii* (AY456250), *Scolecomorphus vittatus* (AY456253), and *Typhlonectes natans* (AF154051).

Molecular and Phylogenetic Analyses

Gene boundaries were determined from sequence data by comparison with other available caecilian mt genomes using MacClade version 4.05 (W. P. Maddison and D. R. Maddison 1992) and PAUP* version 4.0b10 (Swofford 1998).

The phylogenetic relationships of the three *Siphonops* and *Lutkenotyphlus* were inferred using a concatenated data set that included all five tRNA genes of the WANCY region and fragments of the two flanking protein-coding genes (3′-end of the *ND2* gene and 5′-end of *COX1*). Sequences were manually aligned against a previous database (San Mauro et al. 2004), and gaps and ambiguous alignments (42 positions) were excluded from the data using GBLOCKS version 0.91b (Castresana 2000) with default parameters. The final alignment is 572 bp, of which 185 are parsimony informative. The sequences of all other available higher caecilians (*Gegeneophis*, *Scolecomorphus*, and *Typhlonectes*) were used as outgroups. The concatenated alignment was subjected to Bayesian inference (BI; Huelsenbeck et al. 2001), maximum likelihood (ML; Felsenstein 1981), and minimum evolution (ME; Rzhetsky and Nei 1992). All methods were executed using the General Time Reversible (Rodríguez et al. 1990) + Γ model of nucleotide substitution as selected using the Akaike information criterion (Akaike 1973) in Modeltest version 3.6 (Posada and Crandall 1998). BI analysis was conducted with MrBayes version 3.0b4 (Huelsenbeck and Ronquist 2001) simulating four simultaneous chains, for a million generations, sampling every 100 generations, and discarding generations sampled before the chain reached stationarity (100,000) as "burn-in." Statistical support for clades obtained by BI was measured by Bayesian posterior probability. Two independent BI runs were performed to verify congruence of resulting topologies and support. ML and ME analyses were carried out with PAUP*, using heuristic searches with Tree Bisection-Reconnection branch swapping and 10 random stepwise additions of taxa. Support of the resulting ML and ME trees was evaluated by nonparametric bootstrapping with 1,000 pseudoreplicates. The reconstructed phylogeny indicates that *Lutkenotyphlus* is the sister taxon of a monophyletic *Siphonops* (fig. 1). *Siphonops* monophyly is not overwhelmingly robust but receives additional support from the uniquely shared gene order.

To investigate divergence and substitution rates among $tRNA^{Asn}$ genes and pseudogenes found in the three *Siphonops* and *Lutkenotyphlus*, their nucleotide sequences were aligned, together with that of the $tRNA^{Asn}$ gene of *Gegeneophis* (as outgroup), yielding an alignment of 79 bp. Gapped positions were excluded from the alignment, and the resulting 56 sites (33 parsimony informative) were employed to reconstruct a distance phylogeny by ME using JC (Jukes and Cantor 1969) distances (no parameter-richer model was assumed because of the low number of positions analyzed). Relative-rate tests (Robinson et al. 1998) were employed to assess variations in substitution rates using RRTree version 1.1.11 (Robinson-Rechavi and Huchon 2000) assuming JC distances. Base frequencies were compared between $tRNA^{Asn}$ genes and pseudogenes of the three *Siphonops* and *Lutkenotyphlus* using analyses of variance as implemented in STATISTICA version 6.0 (StatSoft Inc. 2001).

The Department of Energy (DOE) Joint Genome Institute database (http://evogen.jgi.doe.gov/) was used to provide comparative information on the 453 complete vertebrate gene mt gene orders included as of April 2005.

Results and Discussion

Rearrangement of the WANCY Region

Our sequencing revealed two different WANCY region gene orders, both of which depart from the consensus order of vertebrates (Seutin et al. 1994; Boore 1999; Jameson et al. 2003) and other analyzed caecilians (Zardoya and Meyer 2000; San Mauro et al. 2004) (fig. 1). The WANCY gene orders of *Siphonops* and *Lutkenotyphlus* are clearly derived. Given that duplications of genes appear to be infrequent among mt genomes (Boore 2000), independent duplications of the WANCY region in

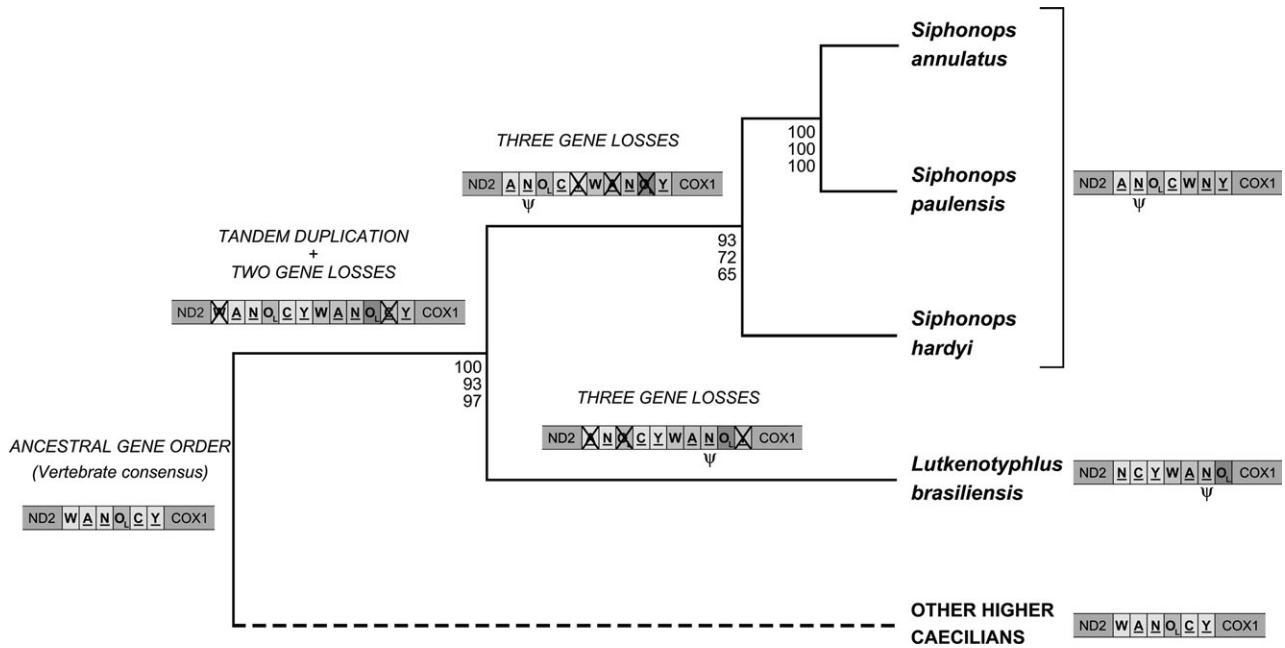


FIG. 1.—Most parsimonious reconstruction of changes producing derived mt gene orders in the WANCY region of three *Siphonops* and *Lutkenotyphlus*. tRNA genes are abbreviated by the corresponding one-letter amino acid code, and genes encoded by the light strand are underlined. ψ indicates the pseudogene. The phylogeny was inferred from a single concatenated data set with all five tRNA genes of the WANCY region and fragments of the two flanking protein-coding genes (*ND2* and *COX1*). Numbers below branches represent support for (from top to bottom) BI, ML, and ME. The derived gene order in *Lutkenotyphlus* dictates that the entire WANCY region must have been involved in the initial tandem duplication, whereas losses reconstructed parsimoniously as occurring before the divergence of *Lutkenotyphlus* and *Siphonops* might plausibly have occurred independently in these lineages.

Lutkentoyphlus and in *Siphonops* provide a less plausible explanation of the derived gene orders of these closely related caecilians than their resulting from a single ancestral tandem duplication of the entire WANCY region followed by almost instant loss of two redundant gene duplicates ($tRNA^{Trp}$, $tRNA^{Cys}$), and independent, random loss of three ($tRNA^{Ala}$, $tRNA^{Asn}$, $tRNA^{Tyr}$) redundant gene duplicates in *Siphonops* and *Lutkenotyphlus* (fig. 1). An alternative reconstruction in which all redundant duplicates are independently lost after the first speciation event (the split between *Siphonops* and *Lutkenotyphlus*) seems equally plausible (not shown).

All rearranged tRNA genes retain the ancestral strand-coding polarity, providing no evidence for inversion. In all *Siphonops* and *Lutkenotyphlus*, there are five intergenic spacers. Most of these range between 4 and 13 nt, and all are in positions expected of pseudogenes under the TDRL model (fig. 2). A more substantial intergenic spacer between the $tRNA^{Ala}$ gene and the O_L is similar to the known, functional $tRNA^{Asn}$ genes of caecilians (fig. 3), but with substantial length and substitution mutations, and can be more confidently identified as the $tRNA^{Asn}$ pseudogene predicted by the TDRL model. All other sequenced caecilian mt genomes (Zardoya and Meyer 2000; San Mauro et al. 2004) typically possess one single intergenic spacer between the WANCY genes, located between $tRNA^{Trp}$ and $tRNA^{Ala}$ (in *T. natans* the spacer is located between $tRNA^{Ala}$ and $tRNA^{Asn}$, and in *S. vittatus* there are no spacers at all between the WANCY genes). In all cases, these spacers comprise a single nucleotide.

Evolution of $tRNA^{Asn}$ Pseudogenes

Although their anticodon sequences are conserved (fig. 3*A*), the *Siphonops* and *Lutkenotyphlus* $tRNA^{Asn}$ pseudogenes have all lost the potential to fold into stable cloverleaf structures, indicating loss of primary function. Moreover, divergence among the pseudogene sequences is far greater than that for their functional paralogs (fig. 3*B*). The phylogeny of all the $tRNA^{Asn}$ paralogs (fig. 3*B*)

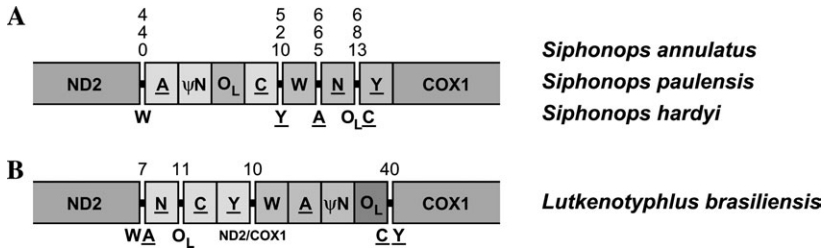


FIG. 2.—Intergenic spacers occurring around the tRNA genes (marked in black) in the WANCY region of the three *Siphonops* (*A*) and *Lutkenotyphlus* (*B*). For every intergenic spacer, its length (in bp, above) in each species and the likely lost gene (below) are shown. tRNA genes are abbreviated by the corresponding one-letter amino acid code, and genes encoded by the light strand are underlined. ψ indicates the pseudogene.

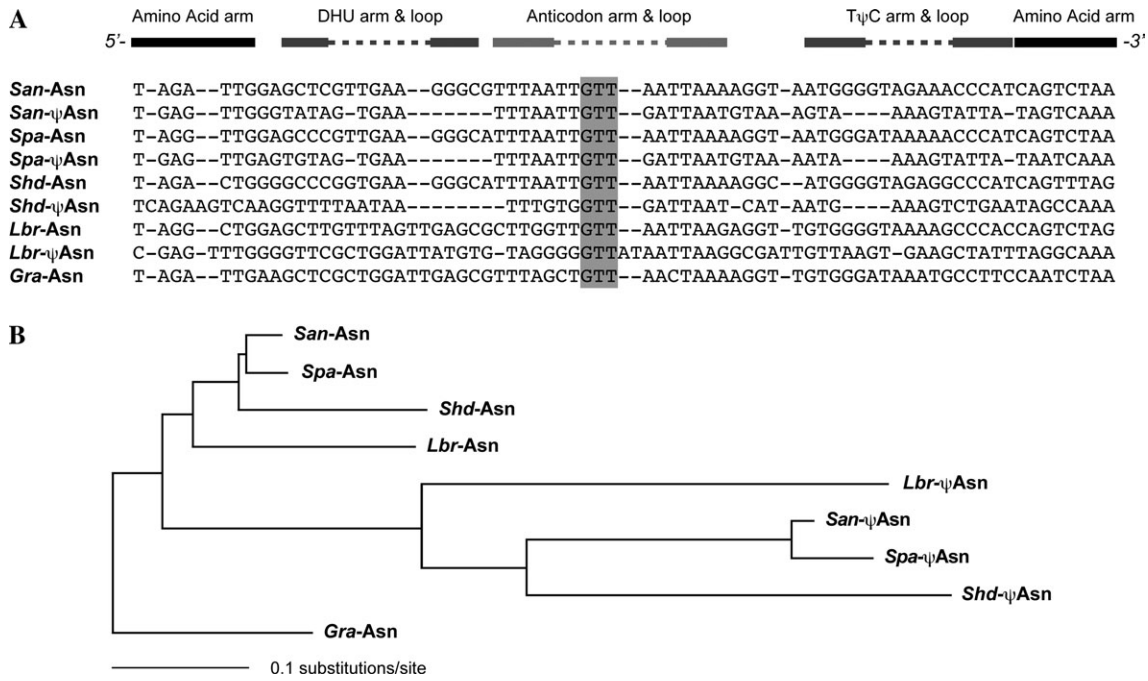


FIG. 3.—Alignment of the $tRNA^{Asn}$ genes and pseudogenes of the three *Siphonops* and *Lutkenotyphlus* (*A*), and ME phylogram inferred from the alignment (after excluding gapped positions) (*B*). tRNA secondary structure is designated above the alignment (full bar indicates arm, dashed bar indicates loop), and the position of the anticodon in highlighted. Species and gene codes are as in table 2. *Gra*, *Gegeneophis ramaswamii* (used as outgroup).

is not as expected because, although their relationships mirror the species phylogeny (fig. 1), duplicates are clustered by functionality instead of homology. Our "incorrect" gene tree probably results from the few data, long branches associated with pseudogenes, and marked base composition differences (see below) between functional and nonfunctional paralogs.

Relative-rate tests show that $tRNA^{Asn}$ pseudogenes evolved more than twice as fast as their functional paralogs (table 2). The contrast of all genes versus all pseudogenes is highly significant, and contrasts of each gene versus its pseudogene paralog are significant with the exception of *Lutkenotyphlus* (table 2). These results suggest that, following duplication, the redundant $tRNA^{Asn}$ paralogs have experienced more relaxed selective constraints (Moritz and Brown 1987). The $tRNA^{Asn}$ pseudogenes have a lower and higher frequency of C and T, respectively (1.7%–9.5% vs. 11.3%–14.3%, $F_{1,6} = 12.760$, $P = 0.012$; 33.3%–38.3% vs. 25.7%–31.5%, $F_{1,6} = 14.086$, $P = 0.010$), than their functional paralogs. Assuming relaxed selection, these biases provide further evidence for asymmetric mutation pressures in mt genomes (Jermiin, Graur, and Crozier 1995).

The pseudogene remnants predicted by TDRL are uncommon in known mt genomes (e.g., Macey et al. 1998; Mueller and Boore 2005; Zhang et al. 2005), consistent with the idea that they are lost rapidly under strong selective pressure to constrain mt genome size and gene number (Wolstenholme 1992). Persistence of an ancestral tandem duplication through a speciation event with subsequent independent random loss of paralogs is a predicted rare event under the TDRL model (Boore 2000) for which our caecilian data may provide the first evidence. Similarly, the $tRNA^{Asn}$ pseudogenes of multiple caecilian lineages provide powerful evidence for TDRL while simultaneously prompting questions about their persistence. $tRNA^{Asn}$ is not distinct from the other four tRNAs in its length and usage, and the $tRNA^{Asn}$ gene is no more or less variable than other caecilian tRNA genes. Their adjacency to O_L is the only obvious variable that correlates with the persistence of these pseudogenes. It may be possible that they (or part of them) have acquired some type of functional role perhaps related to the O_L. Zardoya and Meyer (2000) reported that the O_L of another caecilian, *T. natans*, has the potential to fold into alternative secondary structures with the adjacent $tRNA^{Cys}$. However, similar alternative stem-loop structures have not been found in the caecilian sequences reported here, and the persistence of the pseudogenes is somewhat enigmatic.

Table 2
Results of the Relative-Rate Test for Contrasts Between $tRNA^{Asn}$ Genes and Pseudogenes

Contrast	Rates	SD	*P* Value
All Asn versus all ψAsn	0.320 versus 0.703	0.124	0.002*
San-Asn versus *San*-ψAsn	0.252 versus 0.724	0.165	0.004*
Spa-Asn versus *Spa*-ψAsn	0.278 versus 0.678	0.158	0.011*
Sha-Asn versus *Sha*-ψAsn	0.389 versus 0.772	0.193	0.047*
Lbr-Asn versus *Lbr*-ψAsn	0.360 versus 0.636	0.160	0.085

NOTE.—Results of all possible pairwise contrasts among $tRNA^{Asn}$ genes and among $tRNA^{Asn}$ pseudogenes are nonsignificant ($P > 0.05$). SD, standard deviation; *San*, *Siphonops annulatus*; *Spa*, *Siphonops paulensis*; *Sha*, *Siphonops hardyi*; *Lbr*, *Lutkenotyphlus brasiliensis*; Asn, $tRNA^{Asn}$ gene; and ψAsn, $tRNA^{Asn}$ pseudogene.

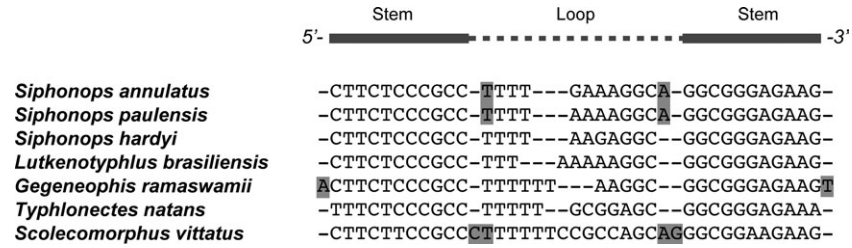


FIG. 4.—Alignment of the O_L's of the three *Siphonops* and *Lutkenotyphlus* (and other higher caecilians). Consensus stem-loop secondary structure is designated above the alignment. Positions highlighted, although outside the indicated consensus stem region, are also part of the stem in those species having them.

Convergence in a Hotspot of Gene Rearrangement

Gene order arrangements may provide exceptionally useful data for phylogenetic inference because of both the relative rarity of rearrangements and the potential complexity of the characters and consequent large character state space (Macey et al. 1997; Boore and Brown 1998; Dowton, Castro, and Austin 2002). These features reduce the chances of homoplasy, and only four convergent derived gene orders among metazoan mitochondria have been previously reported (Flook, Rowell, and Gellissen 1995; Mindell, Sorenson, and Dimcheff 1998; Dowton and Austin 1999; Macey et al. 2004). The derived WANCY region of *Siphonops* is exceptionally similar to that of marsupials (Pääbo et al. 1991), and the order of functional tRNA genes is identical, providing a fifth example of such convergence. The three *Siphonops* differ from marsupials in having a complete O_L in the ancestral vertebrate position relative to the WANCY tRNAs (fig. 1), with nucleotides and stem-loop structures similar to those of other caecilians (fig. 4). This suggests that their O_L's are not secondarily derived, "drifted" duplicated tRNA genes like that found in the derived WANCY region of marsupials (Pääbo et al. 1991). Thus, contrary to previous proposals (Macey et al. 1997, 1998), displacement or loss of the O_L does not always precede vertebrate mt gene order change by tandem duplication.

Tandem duplication can occur during replication by slipped-strand mispairing (Levinson and Gutman 1987) or by illicit priming of replication by tRNAs (Cantatore et al. 1987). In mt genomes, these are thought to particularly involve stem-loop structures and thus to most commonly involve regions including tRNA genes and/or near the origins of replication of the light (O_L in the vertebrate WANCY region) and heavy (O_H in the vertebrate control region) strands (e.g., Moritz and Brown 1987; Pääbo et al. 1991; Stanton et al. 1994; Macey et al. 1997, 1998; Kumazawa et al. 1998; Mindell, Sorenson, and Dimcheff 1998; Boore 1999). Previous studies have cautioned that tandem duplications and gene deletions may be subject to mechanistic constraints such that genes flanking the origins of strand replication are more likely to be duplicated, forming "hotspots" that make convergent gene order rearrangement more probable (Boore and Brown 1998; Mindell, Sorenson, and Dimcheff 1998; Dowton and Austin 1999; Boore 2000).

Ignoring deleted genes and random duplicates, the 453 vertebrate mt genomes in the DOE Joint Genome Institute database display 31 distinct gene orders, with most (368) conforming to the vertebrate consensus. Of the 30 derived gene orders, 4 involve the WANCY region and 26 are found elsewhere in the mt genome. For simplicity, we do not consider rearrangements that involve both the WANCY and other adjacent genomic regions, those evidenced by the genomes of the worm snake *Leptotyphlops dulcis* (Kumazawa and Nishida 1995) and the gluper eels *Eurypharynx pelecanoides* and *Saccopharynx lavenbergi* (Inoue et al. 2003). The four derived gene orders of the WANCY region can be explained by a single TDLR (table 3). Our

Table 3
The 32 Possible Outcomes from Deleting Redundant Gene Copies Subsequent to a Single Tandem Duplication of the Entire Ancestral $W_1A_1N_1C_1Y_1$ Region to Produce $W_1A_1N_1C_1Y_1W_2A_2N_2C_2Y_2$

$W_1A_1N_1C_1Y_1$	Vertebrate consensus
$W_1A_1N_1C_1Y_2$	Vertebrate consensus
$W_1A_1N_1Y_1C_2$	*Chauliodus sloani* (viperfish; Miya, Kawaguchi, and Nishida 2001)
$W_1A_1C_1Y_1N_2$	
$W_1N_1C_1Y_1A_2$	*Hydromantes brunus* (salamander; Mueller et al. 2004)
$A_1N_1C_1Y_1W_2$	
$W_1A_1N_1C_2Y_2$	Vertebrate consensus
$W_1A_1C_1N_2Y_2$	
$W_1A_1Y_1N_2C_2$	*Batrachoseps attenuatus* (salamander; Mueller et al. 2004)
$W_1N_1C_1A_2Y_2$	
$W_1N_1Y_1A_2C_2$[a]	
$W_1C_1Y_1A_2N_2$	
$A_1N_1C_1W_2Y_2$	
$A_1N_1Y_1W_2C_2$[a]	
$A_1C_1Y_1W_2N_2$[a]	
$N_1C_1Y_1W_2A_2$	*Lutkenotyphlus brasiliensis* (caecilian, this study)
$W_1A_1N_2C_2Y_2$	Vertebrate consensus
$W_1N_1A_2C_2Y_2$	
$W_1C_1A_2N_2Y_2$	
$W_1Y_1A_2N_2C_2$	
$A_1N_1W_2C_2Y_2$	
$A_1C_1W_2N_2Y_2$[a]	*Siphonops* species (caecilian, this study; marsupials, Pääbo et al. 1991)
$A_1Y_1W_2N_2C_2$[a]	
$N_1C_1W_2A_2Y_2$	
$N_1Y_1W_2A_2C_2$[a]	
$C_1Y_1W_2A_2N_2$	
$W_1A_2N_2C_2Y_2$	Vertebrate consensus
$A_1W_2N_2C_2Y_2$	
$N_1W_2A_2C_2Y_2$	
$C_1W_2A_2N_2Y_2$	
$Y_1W_2A_2N_2C_2$	
$W_2A_2N_2C_2Y_2$	Vertebrate consensus

[a] A gene order that cannot be explained by a single transposition affected the vertebrate consensus. Although a part of most WANCY regions, the O_L is not considered here. In all cases, it is possible that pseudogenes may remain.

new data for caecilians provide evidence of two derived arrangements of the WANCY region that are also readily explained by the TDRL model of gene order rearrangement. Approximately 15% of all known derived arrangements of the vertebrate mt gene order are explicable in terms of TDRLs of the WANCY region, consistent with the hypothesis that this region may be a mechanistic hotspot of gene duplication by virtue of its association with the O_L.

However, for a tandem duplication to produce a gene order rearrangement, the duplicated region must include more than one complete gene, and the chances of rearrangement are increased with the number of duplicated genes. Thus, tandem duplications are more likely to be detectable in regions, such as the WANCY cluster, with relatively many, small genes, making such regions potential epistemic hotspots. With additional data, it may be possible to address whether rearrangements of the WANCY cluster are significantly more common than expected for any cluster of five small genes and thus better test the hypothesis that the region is a mechanistic hotspot of gene order rearrangement.

Likelihood of Gene Order Change

Dowton, Castro, and Austin (2002) have discussed the probability of convergence in mt gene orders under a "cut and paste" model of gene transposition and inversion. Here, we consider the probability of the observed convergence in the order of tRNAs in the WANCY regions of *Siphonops* and marsupials under the TDRL model. Although which of each of a pair of paralogs is lost or retained subsequent to a single duplication is in principle random (but see Lavrov, Boore, and Brown 2002), those retained from the same duplicate must preserve the original relative order. This leads to some differences in expectations for the TDRL and transposition models. In particular, whereas any derived gene order arrangements that can be explained by a single transposition can also be explained by a single TDRL, some TDRLs produce arrangements that cannot be explained by single transpositions (table 3).

Ignoring changes in the coding strand, there are 120 (5!) possible orders of the five tRNA genes of the WANCY region, suggesting a large character space and low probability of convergence. However, less than a quarter of the arrangements can be produced from the vertebrate consensus by a single TDRL, constraining the character state space and increasing the chance of convergence. There are 32 (2^5) possible random selections of one from each pair of paralogs of a tandemly duplicated WANCY region (not including the O_L) that yield 27 distinct gene orders (table 3). TDRLs of smaller parts of the WANCY region would not add to these 27 different gene orders. Note that six of the 26 derived gene orders cannot be explained by a single transposition (table 3). Note also that six random selections return the original order, so that approximately one-fifth of all WANCY region TDRLs are expected to be undetectable (table 3). In general, for n genes, the probability of undetected TDRLs is $(n + 1)/2^n$. Thus, with fewer genes, the chances of TDRLs being undetected are higher. For example, only one in four TDRLs of two genes yield rearrangements.

The majority, 93 of 119, possible derived gene orders of the vertebrate WANCY region are prohibited by a single TDRL (i.e., require either multiple TDRLs and/or alternative mechanisms of gene order change), but all six currently documented independently derived gene orders found in the WANCY clusters of vertebrates are ones that are permitted by a single TDRL (table 3). Of these six, the convergent WANCY gene orders of marsupial and *Siphonops* cannot be explained by single transpositions, providing further evidence that they have arisen through TDRL. In fact, the conditional probability of at least one convergence given six independent rearrangements produced by single TDRLs of the WANCY region is 0.463 $(1 - ((25/26)\cdot(24/26)\cdot(23/26)\cdot(22/26)\cdot(21/26)))$, so that the observed convergence is hardly surprising given the probable mode of origin.

The Importance of TDRL in Vertebrate mt Evolution

Our data provide compelling evidence, both from the pattern of gene orders and the presence of pseudogenes and intergenic spacers in the positions predicted by the model, that derived caecilian mt gene orders in the WANCY region have evolved through TDRL. Comparing published vertebrate mt gene orders (of 453 complete mt genomes), we find that 24 of the 30 derived arrangements can each be explained by a single TDRL and that the six exceptions can each be explained by two TDRLs. Several of these derived gene orders, like those of *Siphonops* and marsupials, can be explained by a single TDRL but alternatively require multiple transpositions. For example, the highly divergent mt gene order of the gulper eels *E. pelecanoides* and *S. lavenbergi* can be derived from the vertebrate consensus by a single TDRL (Inoue et al. 2003) or by five transpositions. These observations are consistent with the view that TDRL is the dominant mechanism of gene rearrangement in vertebrate mt genomes (e.g., Boore 2000).

Rare genomic changes have attracted great interest because of their potential to provide homoplasy-free evidence of phylogenetic relationships (e.g., Rokas and Holland 2000). Of course, the likelihood of convergence depends on just how rare such changes are, and changes in gene order are not so infrequent that homoplasy is nonexistent (Dowton and Austin 1999; Inoue et al. 2003; Mueller and Boore 2005). The above considerations suggest that convergence in gene order may be more or less common depending also on the mechanism of rearrangement and the mt genomic region considered (Dowton and Austin 1999; Boore 2000; Dowton, Castro, and Austin 2002) and that duplication events may be more or less detectable. In particular, it may be unsurprising if hotspots of tandem duplication coincide with clusters of small genes within which gene order rearrangement is more likely to accompany tandem duplications (Boore 1999). As Darwin (1859) cautioned in *On the Origin of Species*, classifications based on single characters have always failed. Empirical evidence on the relative importance of different mechanisms of gene order rearrangement should provide a basis for more realistic models of gene order rearrangements and best use of comparative gene order data for phylogenetic inference.

Acknowledgments

We are grateful to C. Jared, E. F. Schwartz, C. Schwartz, A. Sebben, and the late C. Zamprogno for invaluable help in obtaining tissue samples. J. A. Cotton and two anonymous reviewers made insightful comments on an earlier version of the manuscript. D.S.M. was sponsored by a predoctoral fellowship of the Ministerio de Educación y Ciencia of Spain and was granted a SYNTHESYS award (Sixth Framework Programme of the European Union) to visit the Natural History Museum. D.J.G. and M.W. were granted BIODIBERIA awards (Fifth Framework Programme of the European Union) to visit the Museo Nacional de Ciencias Naturales. This work received financial support from a project of the Ministerio de Educación y Ciencia of Spain (CGL2004-00401). Fieldwork was supported by National Environmental Research Council GST/02/832.

Literature Cited

Akaike, H. 1973. Information theory as an extension of the maximum likelihood principle. Pp. 267–281 *in* B. N. Petrov and F. Csaki, eds. Second International Symposium of Information Theory. Akademiai Kiado, Budapest, Hungary.

Arndt, A., and M. J. Smith. 1998. Mitochondrial gene rearrangement in the sea cucumber genus *Cucumaria*. Mol. Biol. Evol. **15**:1009–1016.

Blanchette, M., T. Kunisawa, and D. Sankoff. 1999. Gene order breakpoint in animal mitochondrial phylogeny. J. Mol. Evol. **49**:193–203.

Boore, J. L. 1999. Animal mitochondrial genomes. Nucleic Acids Res. **27**:1767–1780.

———. 2000. The duplication/random loss model for gene rearrangement exemplified by mitochondrial genomes of deuterostome animals. Pp. 133–147 *in* D. Sankoff and J. Nadeau, eds. Comparative genomics, computational biology series. Volume 1. Kluwer Academic Publishers, Dordrecht, Netherlands.

Boore, J. L., and W. M. Brown. 1998. Big trees from little genomes: mitochondrial gene order as a phylogenetic tool. Curr. Opin. Genet. Dev. **8**:668–674.

Cantatore, P., M. N. Gadaleta, M. Roberti, C. Saccone, and A. C. Wilson. 1987. Duplication and remoulding of tRNA genes during the evolutionary rearrangement of mitochondrial genomes. Nature **329**:853–855.

Castresana, J. 2000. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. Mol. Biol. Evol. **17**:540–552.

Cosner, M. E., R. K. Jansen, B. M. E. Moret, L.-S. Wang, T. Warnow, and S. Wyman. 2000. An empirical comparison of phylogenetic methods on chloroplast gene order data in Campanulaceae. Pp. 99–121 *in* D. Sankoff and J. Nadeau, eds. Comparative genomics, computational biology series. Volume 1. Kluwer Academic Publishers, Dordrecht, Netherlands.

Darwin, C. 1859. On the origin of species by means of natural selection, or the preservation of favoured races in the struggle for life. John Murray, London.

Dowton, M., and A. D. Austin. 1999. Evolutionary dynamics of a mitochondrial rearrangement "hot spot" in the Hymenoptera. Mol. Biol. Evol. **16**:298–309.

Dowton, M., L. R. Castro, and A. D. Austin. 2002. Mitochondrial gene rearrangements as phylogenetic characters in the invertebrates: the examination of the genome "morphology". Invertebr. Syst. **16**:345–356.

Duellman, W. E., and L. Trueb. 1994. Biology of amphibians. Johns Hopkins University Press, Baltimore, Md.

Felsenstein, J. 1981. Evolutionary trees from DNA sequences: a maximum likelihood approach. J. Mol. Evol. **17**:368–376.

Flook, P., H. Rowell, and G. Gellissen. 1995. Homoplastic rearrangements of insect mitochondrial tRNA genes. Naturwissenschaften **82**:336–337.

Huelsenbeck, J. P., and F. R. Ronquist. 2001. MrBayes: Bayesian inference of phylogeny. Bioinformatics **17**:754–755.

Huelsenbeck, J. P., F. R. Ronquist, R. Nielsen, and J. P. Bollback. 2001. Bayesian inference of phylogeny and its impact on evolutionary biology. Science **294**:2310–2314.

Inoue, J. G., M. Miya, K. Tsukamoto, and M. Nishida. 2003. Evolution of the deep-sea gulper eel mitochondrial genomes: large-scale gene rearrangements originated within the eels. Mol. Biol. Evol. **20**:1917–1924.

Jameson, D., A. P. Gibson, C. Hudelot, and P. G. Higgs. 2003. OGRe: a relational database for comparative analyses of mitochondrial genomes. Nucleic Acids Res. **31**:202–206.

Jermiin, L., D. Graur, and R. H. Crozier. 1995. Evidence from analyses of intergenic regions for strand-specific directional mutation pressure in metazoan mtDNA. Mol. Biol. Evol. **12**:558–563.

Jukes, T. H., and C. R. Cantor. 1969. Evolution of protein molecules. Pp. 21–132 *in* H. N. Munro, ed. Mammalian protein metabolism. Academic Press, New York.

Kumazawa, Y., and M. Nishida. 1995. Variation in mitochondrial tRNA gene organization of reptiles as phylogentic markers. Mol. Biol. Evol. **12**:759–772.

Kumazawa, Y., H. Ota, M. Nishida, and T. Ozawa. 1998. The complete nucleotide sequence of snake (*Dinodon semicarinatus*) mitochondrial genome with two identical control regions. Genetics **150**:313–329.

Larget, B., J. B. Kadane, and D. L. Simon. 2005. A Bayesian approach to the estimation of ancestral genome arrangements. Mol. Phylogenet. Evol. **36**:214–223.

Larget, B., D. L. Simon, J. B. Kadane, and D. Sweet. 2005. A Bayesian analysis of metazoan mitochondrial genome arrangements. Mol. Biol. Evol. **22**:486–495.

Lavrov, D. V., J. L. Boore, and W. M. Brown. 2002. Complete mtDNA sequences of two millipedes suggest a new model for mitochondrial gene rearrangements: duplication and nonrandom loss. Mol. Biol. Evol. **19**:163–169.

Levinson, G., and G. A. Gutman. 1987. Slipped-strand mispairing: a major mechanism for DNA sequence evolution. Mol. Biol. Evol. **4**:203–221.

Liu, Z. Q., Y. Q. Wang, and B. Su. 2005. The mitochondrial genome organization of the rice frog, *Fejervarya limnocharis* (Amphibia: Anura): a new gene order in the vertebrate mtDNA. Gene **346**:145–151.

Lunt, D. H., and B. C. Hyman. 1997. Animal mitochondrial DNA recombination. Nature **387**:247.

Macey, J. R., A. Larson, N. B. Ananjeva, Z. Fang, and T. J. Papenfuss. 1997. Two novel gene orders and the role of light-strand replication in rearrangement of the vertebrate mitochondrial genome. Mol. Biol. Evol. **14**:91–104.

Macey, J. R., T. J. Papenfuss, J. V. Kuehl, H. M. Fourcade, and J. L. Boore. 2004. Phylogenetic relationships among amphisbaenian reptiles based on complete mitochondrial genomic sequences. Mol. Phylogenet. Evol. **33**:22–31.

Macey, J. R., J. A. Schulte II, A. Larson, and T. J. Papenfuss. 1998. Tandem duplication via light-strand synthesis may provide a precursor for mitochondrial genomic rearrangement. Mol. Biol. Evol. **15**:71–75.

Maddison, W. P., and D. R. Maddison. 1992. MacClade: analysis of phylogeny and character evolution. Sinauer Associates Inc., Sunderland, Mass.

Mindell, D. P., M. D. Sorenson, and D. E. Dimcheff. 1998. Multiple independent origins of mitochondrial gene order in birds. Proc. Natl. Acad. Sci. USA **95**:10693–10697.

Miya, M., A. Kawaguchi, and M. Nishida. 2001. Mitogenomic exploration of higher teleostean phylogenies: a case study for moderate-scale evolutionary genomics with 38 newly determined complete mitochondrial DNA sequences. Mol. Biol. Evol. **18**:1993–2009.

Moritz, C., and W. M. Brown. 1986. Tandem duplications of D-loop and ribosomal RNA sequences in lizard mitochondrial DNA. Science **233**:1425–1427.

———. 1987. Tandem duplications in animal mitochondrial DNAs: variation in incidence and gene content among lizards. Proc. Natl. Acad. Sci. USA **84**:7183–7187.

Moritz, C., T. E. Dowling, and W. M. Brown. 1987. Evolution of animal mitochondrial DNA: relevance for population biology and systematics. Annu. Rev. Ecol. Syst. **18**:269–292.

Mueller, R. L., and J. L. Boore. 2005. Molecular mechanisms of extensive mitochondrial gene rearrangement in plethodontid salamanders. Mol. Biol. Evol. **22**:2104–2112.

Mueller, R. L., J. R. Macey, M. Jaekel, D. B. Wake, and J. L. Boore. 2004. Morphological homoplasy, life history evolution, and historical biogeography of plethodontid salamanders inferred from complete mitochondrial genomes. Proc. Natl. Acad. Sci. USA **101**:13820–13825.

Nussbaum, R. A. 1991. Cytotaxonomy of caecilians. Pp. 22–76 *in* S. K. Sessions and D. M. Green, eds. Amphibian cytogenetics and evolution. Academic Press, San Diego, Calif.

Pääbo, S., W. K. Thomas, K. M. Whitfield, and Y. Kumazawa. 1991. Rearrangements of mitochondrial transfer RNA genes in marsupials. J. Mol. Evol. **33**:426–430.

Posada, D., and K. A. Crandall. 1998. MODELTEST: testing the model of DNA substitution. Bioinformatics **14**:817–818.

Robinson, M., M. Gouy, C. Gautier, and D. Mouchiroud. 1998. Sensitivity of the relative-rate test to taxonomic sampling. Mol. Biol. Evol. **15**:1091–1098.

Robinson-Rechavi, M., and D. Huchon. 2000. RRTree: relative-rate tests between groups of sequences on a phylogenetic tree. Bioinformatics **16**:296–297.

Rodríguez, F., J. F. Oliver, A. Marín, and J. R. Medina. 1990. The general stochastic model of nucleotide substitution. J. Theor. Biol. **142**:485–501.

Rokas, A., and P. W. H. Holland. 2000. Rare genomic changes as a tool for phylogenetics. Trends Ecol. Evol. **15**:454–459.

Rzhetsky, A., and M. Nei. 1992. A simple method for estimating and testing minimum-evolution trees. Mol. Biol. Evol. **9**: 945–967.

Sambrook, J., E. F. Fritsch, and T. Maniatis. 1989. Molecular cloning. A laboratory manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.

San Mauro, D., D. J. Gower, O. V. Oommen, M. Wilkinson, and R. Zardoya. 2004. Phylogeny of caecilian amphibians (Gymnophiona) based on complete mitochondrial genomes and nuclear RAG1. Mol. Phylogenet. Evol. **33**:413–427.

San Mauro, D., M. Vences, M. Alcobendas, R. Zardoya, and A. Meyer. 2005. Initial diversification of living amphibians predated the breakup of Pangaea. Am. Nat. **165**:590–599.

Sankoff, D., G. Leduc, N. Antoine, B. Paquin, B. F. Lang, and R. Cedergren. 1992. Gene order comparisons for phylogenetic inference: evolution of the mitochondrial genome. Proc. Natl. Acad. Sci. USA **89**:6575–6579.

Seutin, G., B. F. Lang, D. P. Mindell, and R. Morais. 1994. Evolution of the WANCY region in amniote mitochondrial DNA. Mol. Biol. Evol. **11**:329–340.

Smith, M. J., D. K. Banfield, K. Doteval, S. Gorski, and D. J. Kowbel. 1989. Gene arrangement in sea star mitochondrial DNA demonstrates a major inversion event during echinoderm evolution. Gene **76**:181–185.

Stanton, D. J., L. L. Daehler, C. C. Moritz, and W. M. Brown. 1994. Sequences with the potential to form stem-and-loop structures are associated with coding-region duplications in animal mitochondrial DNA. Genetics **137**:233–241.

StatSoft Inc. 2001. STATISTICA (data analysis software system). (http://www.statsoft.com).

Swofford, D. L. 1998. PAUP*: phylogenetic analysis using parsimony (*and other methods). Version 4.0. Sinauer Associates, Inc., Sunderland, Mass.

Taylor, E. H. 1968. The caecilians of the world: a taxonomic analysis. University of Kansas Press, Lawrence, Kans.

Wilkinson, M. 1997. Characters, congruence and quality: a study of neuroanatomical and traditional data in caecilian phylogeny. Biol. Rev. **72**:423–470.

Wilkinson, M., S. P. Loader, D. J. Gower, J. A. Sheps, and B. L. Cohen. 2003. Phylogenetic relationships of African caecilians (Amphibia: Gymnophiona): insights from mitochondrial rRNA gene sequences. Afr. J. Herpetol. **52**:83–92.

Wilkinson, M., and R. A. Nussbaum. 1992. Taxonomic status of *Pseudosiphonops ptychodermis* Taylor and *Mimosiphonops vermiculatus* Taylor (Amphibia: Gymnophiona: Caeciliaidae) with a description of a new species. J. Nat. Hist. **26**: 675–688.

Wolstenholme, D. R. 1992. Animal mitochondrial DNA: structure and evolution. Int. Rev. Cytol. **141**:173–216.

Zardoya, R., and A. Meyer. 2000. Mitochondrial evidence on the phylogenetic position of caecilians (Amphibia: Gymnophiona). Genetics **155**:765–775.

Zhang, P., H. Zhou, D. Liang, Y. F. Liu, Y. Q. Chen, and L. H. Qu. 2005. The complete mitochondrial genome of a tree frog, *Polypedates megacephalus* (Amphibia: Anura: Rhacophoridae), and a novel gene organization in living amphibians. Gene **346**:133–143.

Franz Lang, Associate Editor

Accepted September 15, 2005

6. PUBLICATION IV

Title: **Experimental design in caecilians systematics: phylogenetic information of mitochondrial genomes and nuclear *rag1***

Authors: Diego San Mauro, James A. Cotton, David J. Gower, Mark Wilkinson, Rafael Zardoya

Status: In preparation

To be submitted to: Systematic Biology (SCI Impact Factor: 10.257)

Resumen IV (Spanish translation of the abstract of Publication IV)

El diseño experimental es un paso obligado de toda investigación científica. En estudios filogenéticos, éste está principalmente relacionado con la elección de genes y taxones que potencialmente mejor responden una cuestión filogenética concreta. Estudios previos han investigado el rendimiento filogenético de diferentes genes y la eficiencia de incrementar el muestreo de taxones, pero sus conclusiones son altamente contradictorias, probablemente porque son muy dependientes del grupo de organismos usados en cada caso. Nosotros exploramos aquí la utilidad filogenética de genes mitocondriales (mt), genomas mt, y el gen nuclear *rag1* para estudios en sistemática de cecilias, así como el efecto que la adición de taxones tienes en la estabilización de un nodo. Primero, se usan métodos basados en verosimilitud para reconstruir una robusta filogenia de cecilias y estimar tiempos de divergencia. Luego, el árbol inferido se usa para cuantificar la información filogenética esperada para diferentes conjuntos de datos y escenarios que puedan ser directamente comparados, proporcionando una valoración objetiva de las estrategias de "adición de secuencias" y "adición de taxones". Se calculan y discuten estimas globales de información por gen, estimas específicas por rama del árbol, estimas de conjuntos de datos combinados (mitogenómicos), y estimas para una rama problemática particular a medida que se van añadiendo taxones hipotéticos en diferentes partes del árbol de las cecilias. En general, los conjuntos de datos más informativos (tanto globalmente como en cada rama) son los de los genes tRNA y ribosomales mt, y algunos genes codificantes de subunidades del complejo NADH deshidrogeneasa (*nad6*, *nad5*, *nad2*). También, el gen nuclear de lenta evolución *rag1* es particularmente informativo en las ramas más internas del árbol. Nuestros resultados también muestran que la adición de taxones en ciertas partes del árbol de las cecilias puede potencialmente incrementar la precisión filogenética sobre la controvertida posición de *Scolecomorphus* y *Boulengerula*. Más ampliamente, la metodología empleada en este estudio permiten una evaluación *a priori* de la conveniencia de un diseño experimental particular para resolver cuestiones específicas a diferentes niveles de la filogenia de las cecilias.

Experimental Design in Caecilians Systematics: Phylogenetic Information of Mitochondrial Genomes and Nuclear *rag1*

DIEGO SAN MAURO[1], JAMES A. COTTON[2], DAVID J. GOWER[3], MARK WILKINSON[3], RAFAEL ZARDOYA[1]

[1]*Departamento de Biodiversidad y Biología Evolutiva, Museo Nacional de Ciencias Naturales - CSIC, José Gutiérrez Abascal 2, 28006 Madrid, Spain.*
[2]*Department of Biology. National University of Ireland, Maynooth, County Kildare, Ireland.*
[3]*Department of Zoology, The Natural History Museum, Cromwell Road, London SW7 5BD, United Kingdom.*

Abstract – Experimental design is a mandatory step of every scientific investigation. In phylogenetic studies, this is mainly related to the choice of genes and taxon sampling that potentially best answers a particular phylogenetic question. While previous studies have investigated the phylogenetic performance of different genes and the effective of increasing taxon sampling, their conclusions are highly contradictory, probably because they are highly dependent on the group of organisms used in each study. We here explore the phylogenetic utility of mitochondrial (mt) genes, mt genomes, and nuclear *rag1* for studies on systematics of caecilian amphibians, as well as the effect of taxon addition on the stabilization of a node. A robust caecilian phylogeny is first reconstructed and divergence times are estimated using likelihood-based methodologies. Then, the inferred tree is used to quantify expected phylogenetic information for different data set scenarios that can be directly compared, providing an objective assessment of "adding sequences" and "adding taxa" strategies. Overall information estimates per gene, specific estimates per branch of the tree, estimates of combined (mitogenomic) data sets, and estimates for a particular controversial branch as hypothetical new taxa are added in different parts of the caecilian tree are calculated and discussed. In general, the most informative data sets (both overall and in most branches) are those for mt tRNA and ribosomal genes, and some mt genes encoding subunits of the NADH dehydrogenase complex (*nad6*, *nad5*, *nad2*). Also, the nuclear slow-evolving *rag1* is particularly informative in the more internal branches of the tree. Our results also show that the addition of taxa in certain parts of the caecilian tree can potentially increase phylogenetic accuracy on the controversial position of *Scolecomorphus* and *Boulengerula*. More broadly, the methodology employed in this study allows an *a priori* evaluation of the appropriateness of particular experimental designs to solve specific questions at different levels of the caecilian phylogeny. [Phylogenetic information; experimental design; taxon sampling; divergence time; mitochondrial genome; mitochondrial genes; *rag1*; Gymnophiona]

The choice of the best genes and taxon sampling for a particular phylogenetic study is usually a hard issue to deal with (Cummings and Meyer, 2005; Graybeal, 1998; Rokas and Carroll, 2005). Given the typically limited resources for phylogenetic studies, it is important to maximize phylogenetic accuracy through an appropriate experimental design. The choice of specific genes and particular taxa that are most appropriate for the phylogenetic question at hand is essential, but also a proper trade-off between adding more taxa versus adding more genes becomes necessary. Several previous studies are largely contradictory on whether it is better to add more genes or more taxa (or, perhaps more importantly, where in a tree is best to add new taxa) to increase phylogenetic accuracy (Graybeal, 1998; Hillis, 1998; Kim, 1996; Kim, 1998; Poe and Swofford, 1999; Pollock and Bruno, 2000; Pollock et al., 2002; Rannala et al., 1998; Rokas and Carroll, 2005; Rosenberg and Kumar, 2001; Zwickl and Hillis, 2002).

In general, the use of "favourite" genes or genomic regions in phylogenetic studies is more commonly related to the technical ease with which their sequences can be determined (e.g., availability of primers) and their "success" in previous similar-level studies, rather than to quantitative results on their actual appropriateness for a particular phylogenetic question (Cummings and Meyer, 2005). Nevertheless, several empirical studies have previously investigated the phylogenetic performance of individual genes in reconstructing a given phylogeny (considered as "true", and usually derived from a much larger data set) under different inference frameworks. The vast majority of these studies evaluated the performance of mitochondrial (mt) genes, and used the mitogenomic (Curole and Kocher, 1999) tree as the reference topology (Cummings et al., 1995; Miya and Nishida, 2000; Mueller, 2006; Russo et al., 1996; Zardoya and Meyer, 1996). A few others compared, either directly or indirectly, the utility of nuclear versus mt genes (Graybeal, 1994; Groth and Barrowclough, 1999; San Mauro et al., 2004b; Springer et al., 2001), or used simulations to explore how rates of molecular evolution influence phylogenetic reconstruction (Yang, 1998). In addition, many of these studies indicated that rather large sequence data sets are needed to achieve statistical confidence in phylogenetic inference, and suggested the combination of several genes to tackle high-level phylogenetic questions (Cummings et al., 1995; Russo et al., 1996; Zardoya and Meyer, 1996).

While several general conclusions can be drawn from these previous studies, such as e.g. the high performance of mt ribosomal genes or the low performance of *nad4L*, the fact is that there are many discrepant results among them, which essentially precludes a broad generalization of their conclusions. Overlooking differences in employed phylogenetic inference methods, those discrepancies may be related to the fact that each study is based, at least, on different taxon samplings, and, in most cases, on completely different groups of organisms. Thus, each study is evaluating markedly different data sets, and the generality of their particular results might be compromised beyond the group of organisms used in the study (Russo et al., 1996).

We recently determined the complete mt genome and partial nuclear *rag1* sequences of several caecilians amphibians (order Gymnophiona), and used them to infer phylogenetic relationships of major families within the group (San Mauro et al., 2004b). That study revealed the utility of those two molecular markers, but also prompted some questions on experimental design for future studies on caecilian systematics at different levels. Caecilians are elongate, limbless tropical amphibians that, albeit being a relatively small group (about 170 currently recognized species; AmphibiaWeb, 2006; Frost, 2004), possess a remarkable morphological (Taylor, 1968; Wilkinson and

Nussbaum, 1997), ecological (Gower et al., 2004; Loader et al., 2003), reproductive (Kupfer et al., 2006; Wake, 1977), and developmental (Müller, 2006; Müller et al., 2005) diversity that, only recently, is becoming known. Therefore, it is getting more and more necessary to reconstruct a robust phylogenetic framework that will allow a proper exploration of the evolutionary implications arising from coming caecilian studies.

The aim of this study is to explore the phylogenetic utility of different data sets of mt genes, the mt genome, and nuclear *rag1* for future experimental designs on caecilian systematics at different divergence levels, and to assess where in the phylogeny new caecilian taxa should be added to increase phylogenetic accuracy in specific parts of the tree. Unlike previous studies, our approach does not evaluate the performance of each gene in reconstructing a given phylogeny, but rather estimates what information is expected to be in a particular data set, and how the expected information about a particular controversial edge varies with addition of hypothetical taxa, using the method of Goldman (1998). This method allows information scores for different data set scenarios to be straightforwardly compared, thus allowing objective comparisons of alternative "adding sequences" and "adding taxa" strategies.

MATERIALS AND METHODS

Taxon Sampling and DNA Sequencing

Our study includes representatives of nine genera of caecilian amphibians, covering all six currently recognized families (Table 1). San Mauro et al. (2004b) indicated that Caeciliidae (the most diverse and cosmopolitan caecilian family; Nussbaum and Wilkinson, 1989; Taylor, 1968; Wilkinson and Nussbaum, 2006) was inadequately represented with a single representative due to its paraphyletic condition with respect to the aquatic Typhlonectidae (Frost et al., 2006; Hedges et al., 1993; Nussbaum, 1979; Roelants et al., in ms.; Wilkinson, 1997; Wilkinson et al., 2003), and perhaps to also Scolecomorphidae (Frost et al., 2006; Wilkinson et al., 2003). Thus, in addition to the Indian *Gegeneophis ramaswamii* used in San Mauro et al. (2004b), we have included three other species representing major, distinct lineages within the Caeciliidae: the East African *Boulengerula taitanus*, the West African *Geotrypetes seraphini*, and the South American *Siphonops annulatus* (Wilkinson, 1997).

The nucleotide sequences of the complete mt genomes of *B. taitanus* and *G. seraphini* were determined anew for this study. Also, a 1,509 base pair (bp) long fragment of the nuclear *rag1* was determined in *B. taitanus*, *G. seraphini*, and *S. annulatus*. In all cases, total DNA was purified from ethanol-preserved liver with standard phenol/ chloroform extraction procedures (Sambrook et al., 1989), and nucleotide sequences were determined using the primers, conditions, and methods reported in San Mauro et al. (2004b). Mt genome and *rag1* sequence information was available for all other employed caecilian taxa from previous studies (San Mauro et al., 2004b; San Mauro et al., 2006; Zardoya and Meyer, 2000). Details of the employed species, taxonomic assignment, voucher specimens, collection localities, and GenBank accession numbers can be found in Table 1. Distinct structural features of the mt genomes of *B. taitanus* and *G. seraphini* are presented in the Appendix.

Sequence Alignments and Phylogenetic Reconstruction

Alignments were separately prepared for each mt and nuclear *rag1* data set. Nucleotide sequences of mt *rrnS* and *rrnL* genes were aligned using Clustal X version 1.83 (Thompson et al., 1997), and revised by eye in order to maximize positional homology. Sequences of each mt tRNA gene (except *trnF* that is lacking in *G. ramaswamii*'s mt genome; San Mauro et al., 2004b) were manually aligned based on the corresponding cloverleaf secondary structure, and then combined into a single concatenated data set. Deduced amino acid sequences of all 13 mt protein-coding genes were manually aligned against a previous database (San Mauro et al., 2004b), and the resulting alignments were then imposed onto the corresponding nucleotide sequences (used in all subsequent analyses). In all cases, gaps and alignment ambiguities were excluded from the data sets using GBlocks version 0.91b (Castresana, 2000) with default parameters. *Rag1* nucleotide sequences were manually aligned against San Mauro et al.'s (2004b) database.

All data sets (mt ribosomal, tRNA, and protein-coding genes, and nuclear *rag1*) were combined to reconstruct a caecilian phylogeny based in the largest and most comprehensive set of sequence characters available. Third codon positions of mt protein-coding genes were excluded from the phylogenetic analyses

TABLE 1. Data for caecilian samples employed in this study.

Species	Family	Voucher number	Collection locality	GenBank accession nos. (mt genomes, *rag1*)
Rhinatrema bivittatum	Rhinatrematidae	BMNH 2002.6	Kaw, French Guyana	AY456252, AY456257
Ichthyophis glutinosus	Ichthyophiidae	MW 1733	Peradeniya, Sri Lanka	AY456251, AY456256
Uraeotyphlus oxyurus	Uraeotyphlidae	MW 212	Payyanur, India	AY456254, AY456259
Scolecomorphus vittatus	Scolecomorphidae	BMNH 2002.100	Amani, Tanzania	AY456253, AY456258
Typhlonectes natans	Typhlonectidae	BMNH 2000.218[a]	Potrerito, Venezuela[a]	AF154051, AY456260
Gegeneophis ramaswamii	Caeciliidae	MW 331	Thenmalai, India	AY456250, AY456255
Boulengerula taitanus	Caeciliidae	NMK A/3112	Wundanyi, Kenya	AY954504[b], DQ320062[b]
Geotrypetes seraphini	Caeciliidae	BMNH 2005.2	Cameroon (no locality – pet trade)	AY954505[b], DQ320063[b]
Siphonops annulatus	Caeciliidae	BMNH 2005.9	Dominguez Martins, Brazil	AY954506, DQ320064[b]

BMNH, The Natural History Museum, London (UK).
MW, field series of the Zoology Department, University of Kerala (India) and the Department of National Museums, Colombo (Sri Lanka).
NMK, National Museums of Kenya, Nairobi (Kenya).
[a]Only for the specimen used to sequence *rag1*. Source data for the specimen used to sequence the mt genome are unknown (pet trade).
[b]Determined for this study.

because transitions were saturated, as judged by plots of pairwise uncorrected (transition and transversion) differences versus corrected sequence divergence (measured as maximum likelihood distance) (not shown). All our phylogenetic trees were rooted on the rhinatrematid caecilian *Rhinatrema bivittatum*. This rooting is justified by previous molecular (Frost et al., 2006; Hedges et al., 1993; Roelants et al., in ms.; San Mauro et al., 2004b; San Mauro et al., 2005) and morphological (Nussbaum, 1977; Nussbaum, 1979; Wilkinson, 1992; Wilkinson, 1996; Wilkinson, 1997) studies that have supported that the Rhinatrematidae is the sister group of all other extant caecilians.

The combined data set was subjected to maximum likelihood (ML; Felsenstein, 1981) and Bayesian inference (BI; Huelsenbeck et al., 2001) methods of phylogenetic reconstruction. ML analysis was performed with PAUP* version 4.0b10 (Swofford, 1998), using heuristic searches with ten random stepwise additions of taxa and Tree Bisection and Reconnection branch swapping. BI analysis was conducted with MrBayes version 3.1.2 (Huelsenbeck and Ronquist, 2001; Ronquist and Huelsenbeck, 2003) running four simultaneous Markov chains, for five million generations, sampling every 100 generations, and discarding generations sampled before the chain reached stationarity (one million) as burn-in. Two independent BI runs were performed to control for an adequate mixing of the Markov chains, and to verify congruence of resulting topologies, parameters, and support.

For both ML and BI, best-fit models of nucleotide substitution were selected using Modeltest version 3.7 (Posada and Crandall, 1998) following the Akaike information criterion (AIC; Akaike, 1973). For ML, one single model of nucleotide substitution was selected: the General Time Reversible (GTR; Rodriguez et al., 1990) with gamma-distributed among-site rate heterogeneity approximated with four categories (Γ_4; Yang, 1994) and a proportion of invariant sites (I; Reeves, 1992). For BI, independent best-fit models were selected for each of the following seven partitions: first codon positions of mt protein-coding genes (GTR+Γ_4+I), second codon positions of mt protein-coding genes (GTR+Γ_4+I), first codon positions of *rag1* (GTR+I), second codon positions of *rag1* (GTR+I), third codon positions of *rag1* (GTR+Γ_4), mt ribosomal genes (GTR+Γ_4+I), and mt tRNA genes (GTR+Γ_4).

Support of the resulting ML tree was evaluated by non-parametric bootstrapping with 2,000 pseudo-replicates. Statistical support for clades obtained by BI was measured by Bayesian posterior probability.

Evaluation of Alternative Tree Topologies

Five alternative tree topologies (see results) were evaluated by parametric bootstrapping (PB; Efron, 1985; Goldman, 1993; Huelsenbeck et al., 1996a) and non-parametric approximately unbiased (AU; Shimodaira, 2002) test.

PB was conducted using HyPhy version 0.99b (Kosakovsky Pond et al., 2005) simulating series of 1,000 replicate data sets under seven independent GTR+Γ_4 models, assigned to the same partitions mentioned for the BI analysis. A Bonferroni multiple-test correction (Bonferroni, 1936) was applied to adjust the level of significance for rejection of alternative hypotheses.

AU test was carried out using Consel version 0.1i (Shimodaira and Hasegawa, 2001) with site-wise log-likelihoods calculated by PAML version 3.15 (Yang, 1997) with independent GTR+Γ_4 models assigned to the same partitions indicated for BI. A total of one million multiscale bootstrap replicates were used in order to reduce sampling error.

Estimation of Divergence Times

Non-constancy of rates of evolution among lineages was determined using a likelihood ratio test statistic (LRT; Felsenstein, 1981; Huelsenbeck and Crandall, 1997) calculated for the ML tree under the GTR+Γ_4+I model, with and without enforcing the molecular clock (as implemented in PAUP*).

Divergence times were determined using a Bayesian relaxed molecular clock approach that allows variation of rates of evolution among genes and among lineages, and uses a Markov chain Monte Carlo method to approximate both prior and posterior distributions of divergence time estimates (Kishino et al., 2001; Thorne and Kishino, 2002; Thorne et al., 1998). We used the best topology that was inferred from both ML and BI analyses (same in both cases, see results) as the starting phylogeny. The sequences of a frog (*Alytes obstetricans*: mt genome AY585337 [San Mauro et al., 2004a], *rag1* AY583334 [San Mauro et al., 2005]) and a salamander (*Lyciasalamandra atifi*: mt genome AF154053 [Zardoya et al., 2003], *rag1* AY456261 [San Mauro et al., 2004b]) were used as outgroups for this analysis. Branch lengths of the inferred topology, and divergence times were estimated using PAML and the programs Estbranches and Multidivtime (available at http://statgen.ncsu.edu/thorne/multidivtime.html). The Markov chain was run for ten million generations, with sampling intervals of 100 generations, and a burn-in period corresponding to the first million generations.

The prior assumption for the mean of the time of the ingroup root node (rttm) was set to 3.32 time units, where 1 time unit in this analysis represents 100 million years (Myr). This value was obtained based on the mean of the estimated split of caecilians and batrachian (frogs + salamanders) lineages 332 Myr ago (Mya) (Roelants et al., in ms.). The standard deviation of the prior distribution was set to its maximum value (equal to the mean) to avoid violation of the definition of a prior. Given that the fossil record of caecilians is so far insufficient to provide any reliable basis for calibration, we calibrated our relaxed clock using age estimates from a recent molecular-based study (Roelants et al., in ms.). We used six time constraints on three internal nodes (see phylogenetic results below): (1) split between the ichthyophiid + uraeotyphlid lineage and the higher caecilians between 167 and 132 Mya, (2) split between ichthyophiid and uraeotyphlid caecilians between 70 and 44 Mya, (3) split between *Gegeneophis* and the *Geotrypetes* + *Siphonops* clade between 101 and 81 Mya.

TABLE 2. Best-fitting substitution models (and associate estimated parameters) used for assessing phylogenetic information content of each employed data set.

Data set name (genes included, and number of positions in parentheses)	Best-fit model	Base frequencies		Substitution rate matrix (Ti:Tv ratio for HKY models)		Γ_4-shape parameter (α)	Prop. invariant sites (I)
AT6 (*atp6* without the 3rd position; 452 bp)	TrN+Γ_4	A = 0.239 C = 0.296	G = 0.114 T = 0.350	A-C = 1.000 A-G = 2.733 A-T = 1.000	C-G = 1.000 C-T = 4.208 G-T = 1.000	0.296	0
AT8 (*atp8* without the 3rd position; 42 bp)	HKY+Γ_4	A = 0.234 C = 0.231	G = 0.090 T = 0.445	1.642		0.311	0
CO1 (*cox1* without the 3rd position; 1,020 bp)	TrN+Γ_4+I	A = 0.229 C = 0.235	G = 0.215 T = 0.321	A-C = 1.000 A-G = 4.196 A-T = 1.000	C-G = 1.000 C-T = 10.854 G-T = 1.000	1.401	0.747
CO2 (*cox2* without the 3rd position; 450 bp)	TVM+Γ_4+I	A = 0.299 C = 0.229	G = 0.171 T = 0.302	A-C = 1.300 A-G = 6.030 A-T = 1.924	C-G = 0.212 C-T = 6.030 G-T = 1.000	1.950	0.593
CO3 (*cox3* without the 3rd position; 522 bp)	TVM+Γ_4+I	A = 0.216 C = 0.246	G = 0.211 T = 0.327	A-C = 4.595 A-G = 13.714 A-T = 4.600	C-G = 0.673 C-T = 13.714 G-T = 1.000	1.126	0.640
CYB (*cob* without the 3rd position; 752 bp)	GTR+Γ_4+I	A = 0.256 C = 0.250	G = 0.167 T = 0.327	A-C = 2.698 A-G = 3.428 A-T = 1.436	C-G = 0.566 C-T = 8.745 G-T = 1.000	0.716	0.579
ND1 (*nad1* without the 3rd position; 612 bp)	GTR+Γ_4+I	A = 0.232 C = 0.281	G = 0.169 T = 0.318	A-C = 0.402 A-G = 2.124 A-T = 0.854	C-G = 0.263 C-T = 3.226 G-T = 1.000	0.630	0.487
ND2 (*nad2* without the 3rd position; 678 bp)	TVM+Γ_4+I	A = 0.279 C = 0.280	G = 0.115 T = 0.326	A-C = 1.644 A-G = 4.346 A-T = 1.411	C-G = 0.453 C-T = 4.346 G-T = 1.000	0.885	0.313
ND3 (*nad3* without the 3rd position; 216 bp)	TVM+Γ_4	A = 0.206 C = 0.295	G = 0.156 T = 0.344	A-C = 31.609 A-G = 95.307 A-T = 28.624	C-G = 6.870 C-T = 95.307 G-T = 1.000	0.279	0
ND4 (*nad4* without the 3rd position; 900 bp)	GTR+Γ_4+I	A = 0.275 C = 0.264	G = 0.135 T = 0.327	A-C = 1.779 A-G = 3.678 A-T = 1.560	C-G = 0.470 C-T = 6.716 G-T = 1.000	1.012	0.413
ND4L (*nad4L* without the 3rd position; 184 bp)	GTR+Γ_4+I	A = 0.222 C = 0.269	G = 0.169 T = 0.340	A-C = 1.111 A-G = 2.608 A-T = 2.360	C-G = 0.614 C-T = 5.151 G-T = 1.000	2.695	0.461
ND5 (*nad5* without the 3rd position; 1,098 bp)	GTR+Γ_4+I	A = 0.289 C = 0.246	G = 0.155 T = 0.311	A-C = 3.100 A-G = 4.051 A-T = 2.170	C-G = 0.706 C-T = 9.505 G-T = 1.000	1.071	0.429
ND6 (*nad6* without the 3rd position; 286 bp)	GTR+Γ_4	A = 0.134 C = 0.142	G = 0.280 T = 0.444	A-C = 0.820 A-G = 6.373 A-T = 2.806	C-G = 1.172 C-T = 2.365 G-T = 1.000	0.538	0
PROTS-NO3 (mt protein-coding genes without the 3rd position; 7,212 bp)	GTR+Γ_4+I	A = 0.249 C = 0.253	G = 0.170 T = 0.328	A-C = 1.889 A-G = 3.594 A-T = 1.800	C-G = 0.493 C-T = 6.515 G-T = 1.000	0.914	0.487
PROTS-ALL (mt protein-coding genes - all positions; 10,818 bp)	GTR+Γ_4+I	A = 0.338 C = 0.242	G = 0.125 T = 0.295	A-C = 1.501 A-G = 3.105 A-T = 2.150	C-G = 0.306 C-T = 11.817 G-T = 1.000	1.022	0.355
3rdPOS (3rd positions of mt protein-coding genes; 3,606 bp)	GTR+Γ_4+I	A = 0.469 C = 0.210	G = 0.059 T = 0.262	A-C = 0.000 A-G = 8.826 A-T = 0.060	C-G = 0.000 C-T = 11.456 G-T = 1.000	0.352	0.007
12S (mt *rrnS*; 699 bp)	GTR+Γ_4	A = 0.327 C = 0.245	G = 0.198 T = 0.231	A-C = 7.324 A-G = 19.659 A-T = 12.774	C-G = 0.000 C-T = 37.794 G-T = 1.000	0.510	0
16S (mt *rrnL*; 1,169 bp)	GTR+Γ_4+I	A = 0.379 C = 0.207	G = 0.175 T = 0.240	A-C = 3.878 A-G = 5.883 A-T = 5.736	C-G = 0.000 C-T = 19.712 G-T = 1.000	0.883	0.306
tRNAs (all mt tRNA genes except *trnF*; 1,278 bp)	TVM+Γ_4	A = 0.317 C = 0.183	G = 0.182 T = 0.318	A-C = 1.673 A-G = 10.638 A-T = 2.136	C-G = 0.000 C-T = 10.638 G-T = 1.000	0.800	0
mtGENOME-NO3 (all single mt data sets combined, excluding 3rd positions; 10,358bp)	GTR+Γ_4+I	A = 0.278 C = 0.237	G = 0.174 T = 0.311	A-C = 2.182 A-G = 5.165 A-T = 2.333	C-G = 0.366 C-T = 8.326 G-T = 1.000	1.031	0.422
RAG1 (nuclear *rag1*; 1,509 bp)	GTR+Γ_4	A = 0.311 C = 0.199	G = 0.227 T = 0.263	A-C = 1.748 A-G = 6.635 A-T = 1.476	C-G = 0.970 C-T = 9.758 G-T = 1.000	0.457	0

Estimation of Phylogenetic Information

Phylogenetic information was estimated using the method described by Goldman (1998). The method allows quantifying expected likelihood information scores (derived from the Fisher information matrix; Atkinson and Donev, 1992; Edwards, 1972) in a sequence alignment based on parameters of a Markov model of nucleotide substitution for a particular phylogenetic tree.

Twenty mt data sets and one nuclear *rag1* data set were prepared, and best-fit models of nucleotide substitution for each one were selected following the AIC in Modeltest. Genes included in each data set, as well as their lengths, best-fit models, and associate parameters are shown in Table 2. The program Edible (Massingham and Goldman, 2000) was employed to estimate expected phylogenetic information using model parameters of each data set, and the best topology inferred from both ML and BI analyses (same in both cases, see results). Phylogenetic information is quantified per site. In order to obtain estimates per data set for the overall comparisons, the per site scores were multiplied up by the partition length. To make phylogenetic information calculations comparable among the different data sets, the base trees need to be scaled to the same overall rate variation, which entails assuming not only the same topology, but also the same branch lengths on them. We used the branch lengths optimized for the full data set employed in the phylogenetic reconstruction analyses (i.e., mt ribosomal, tRNA, and protein-coding genes, and nuclear *rag1* combined), as it encompasses rate variation of all source genes. In addition, we estimated phylogenetic information scores using the branch lengths optimized for the RAG1 data set (only-nuclear subset), the mtGENOME-NO3 data set (only-mt subset), and the full data set using 16 categories (instead of four) to approximate the gamma-shape parameter.

Phylogenetic information scores were also estimated per branch of the caecilian unrooted tree for each of the single data sets. In this case, the base tree was again scaled to the overall rate variation of the full data set.

Goldman's (1998) method also allows assessing the variation in phylogenetic information for particular branches when adding new hypothetical taxa in different parts of the tree, that is, it quantifies the information about the branch of interest as taxa are added in different other branches, and at different positions along each branch. In our study, we used this approach to identifying branches where is best to add new taxa in order to increase phylogenetic information of a particular controversial branch of the caecilian phylogeny. The base tree was again scaled to the overall rate variation of the full data set (as it is the most comprehensive), and hypothetical new taxa were separately added to all nine terminal branches and five internal branches, at twelve (evenly distributed) different positions along each branch. For terminal edges, we made each newly added branch to be the same length as the edge it was being added next to, and, for internal edges, the new braches were added as the mean of the descendents, so the tree was as nearly ultrametric as possible.

Some statistical tests (analysis of variance, analysis of covariance, linear regression) were conducted using STATISTICA version 6.0 (StatSoft Inc., 2001).

RESULTS AND DISCUSSION

Caecilian Phylogeny

The final combined alignment was 11,867 bp (after exclusion of gaps, alignment ambiguities, and third codon positions of mt protein-coding genes), of which 7,221 were invariant, and 2,683 were parsimony-informative. Both ML (-ln L = 57,002.416) and BI (-ln L = 55,825.980 for run 1; -ln L = 55,829.220 for run 2) analyses yielded the same inferred relationships among caecilian taxa with differences only in branch lengths and levels of support (Fig. 1). ML bootstrap support is substantial (>75) to maximal for all nodes except one (node D) that only receives weak support. Bayesian posterior probabilities are maximal or nearly so for all nodes.

Outside the higher caecilians, the recovered tree is in full agreement with the most recent molecular (Frost et al., 2006; Roelants et al., in ms.; San Mauro et al., 2004b; San Mauro et al., 2005; Wilkinson et al., 2003; Wilkinson et al., 2002) and morphological (Wilkinson, 1997; Wilkinson and Nussbaum, 1996; Wilkinson and Nussbaum, 1999) studies that support the sister group relationship of Ichthyophiidae and Uraeotyphlidae, and the monophyly of higher caecilians (Wilkinson and Nussbaum, 2006). Unlike this general agreement outside the higher caecilians, there is a lot more uncertainty on the inter- and intrafamilial phylogenetic relationships within the higher caecilians (San Mauro et al., 2004b; Wilkinson, 1997; Wilkinson et al., 2003; Wilkinson and Nussbaum, 2006). Indeed, the most comprehensive morphological study to date (Wilkinson, 1997) was unable to conclusively resolve relationships among these higher caecilians, and only two recent molecular studies (Frost et al., 2006; Roelants et al., in ms.) have provided a supported (but conflicting with each other) picture of the phylogenetic relationships within this clade. The differences between these two molecular studies may rely on the methodological approach employed to reconstruct phylogenetic relationships: parsimony-based in Frost et al.'s (2006) study versus likelihood-based in Roelants et al.'s (in ms.). Our phylogenetic results are in full agreement with those by Roelants et al. (in ms.) that reflect a perhaps more traditional assemblage, with Scolecomorphidae recovered as the sister group of all other higher caecilians (Duellman and Trueb, 1994; Nussbaum and Wilkinson, 1989), and with Caeciliidae recovered as paraphyletic with respect to Typhlonectidae (Hedges et al., 1993; Nussbaum, 1979; Wilkinson, 1997; Wilkinson et al., 2003). Despite both Roelants et al.'s (in ms.) and our phylogenetic results are likelihood-based, the complete congruence among them is not trivial given the markedly differences between the employed data sets: Roelants et al.'s (in ms.) being more nuclear-based (one mt gene fragment [14%] + four nuclear gene fragments [86%]) and including representatives of all amphibian lineages (171 amphibian taxa, of which 24 are caecilians) and some amniote outgroups, and ours being more mt-based (complete mt genome [87%] + one nuclear gene fragment [13%]) and using exclusively (nine) caecilian lineages.

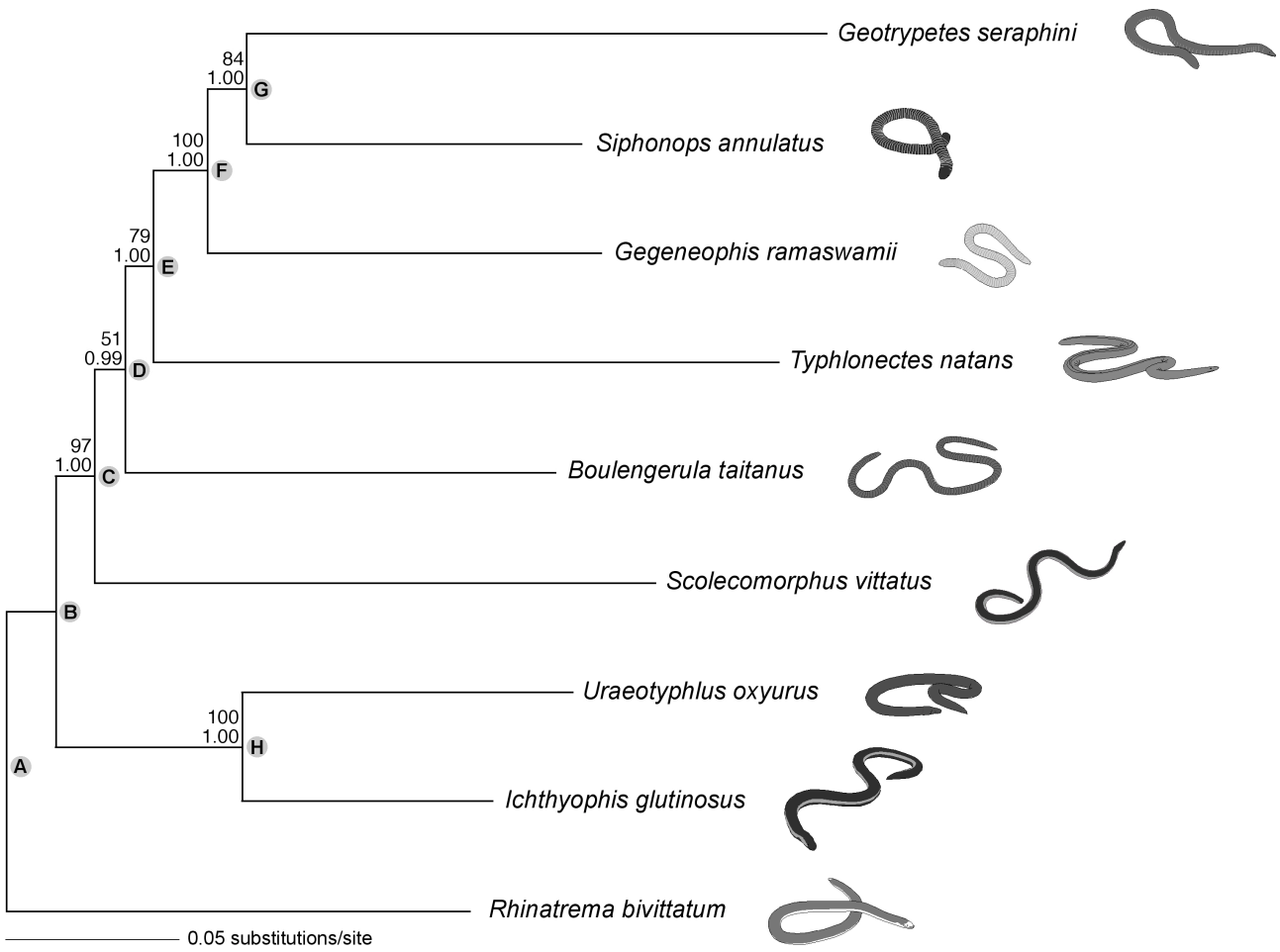


FIGURE 1. Maximum likelihood phylogram of the studied caecilian lineages inferred from a combined data set of mt rRNA, tRNA, and protein-coding genes, and nuclear *rag1* (see text). Numbers above branches represent support for ML (bootstrap proportions; upper value), and BI (posterior probabilities; lower value). Letters next to nodes refer to those of Table 4 for divergence time estimates.

In contrast to our and Roelants et al.'s (in ms.) results, Frost et al.'s (2006) study indicated that Caeciliidae is paraphyletic with respect to Typhlonectidae as well as with respect to Scolecomorphidae. This hypothesis had already been suggested (albeit only tentatively) by a previous molecular study (Wilkinson et al., 2003), and involves an alternative branching scenario at the base of the higher caecilian tree (with a *Boulengerula* + *Herpele* clade recovered as the sister group of all other higher caecilians). Interestingly, the only node that is not confidently supported by the two methods of phylogenetic inference used in our study (receiving nearly maximal support by BI, but only weak by ML) is the one of the basal split within the higher caecilians (node D in Figure 1), involving the phylogenetic position of *Scolecomorphus* and *Boulengerula*. We used PB and AU test to evaluate the three alternative branching arrangements of these two caecilians at the base of the higher caecilian tree: *Scolecomorphus* as the sister taxon of all other higher caecilians (unconstrained), *Boulengerula* as the sister taxon of all other higher caecilians, *Scolecomorphus* + *Boulengerula* as the sister clade of all other higher caecilians. In addition, the full caecilian trees recovered by Wilkinson et al. (2003) and Frost et al. (2006) were also evaluated. The results of the two employed tests are summarized in Table 3. PB strongly rejects all constrained topologies (at Bonferroni-corrected α = 0.0125), whereas the AU test only allows rejection (at α = 0.05) of the topologies by Wilkinson et al. (2003) and Frost et al. (2006) (topologies 4 and 5 in Table 3). There seems to be a correlation between AU rejection and ML bootstrap support, with those trees constraining nodes with substantial ML bootstrap values (those by Wilkinson et al. [2003] and Frost et al. [2006]) being rejected by the AU test.

It has been proposed that the observed discrepancy between results from parametric and non-parametric likelihood-based tests may be related to different forms of null hypothesis: while parametric tests evaluate that a given topology is the true topology, non-parametric tests evaluate whether two (or more) topologies are equally good explanations of the data (Buckley, 2002; Goldman et al., 2000). Parametric tests are usually considered more powerful than non-parametric tests (Goldman et al., 2000), but also too liberal, with a higher Type 1 error rate derived from the use of oversimplified models of sequence evolution to construct the null distribution (Buckley, 2002; Goldman et al., 2000; Strimmer and Rambaut, 2001). On this point, Huelsenbeck et al. (1996b) note that parametric tests should be implemented with highly complex models of sequence evolution to bring Type 1 errors to the lowest possible. We have attempted to reduce model misspecification in our PB analysis by using independent complex, parameter-rich models of sequence evolution assigned to different partitions of

TABLE 3. Log-likelihoods and *P*-values of parametric bootstrapping (PB) and approximately unbiased (AU) test for each of the alternative topologies evaluated.

Alternative topologies	-ln *L*[a]	*P* (PB)	*P* (AU)
1. (*Rbi*,((*Igl*,*Uox*),(*Svi*,(*Bta*,(*Tna*,(*Gra*,(*San*,*Gse*)))))))[b]	55,458.032	–	0.636
2. (*Rbi*,((*Igl*,*Uox*),(*Bta*,(*Svi*,(*Tna*,(*Gra*,(*San*,*Gse*)))))))	55,460.197	0.001	0.505
3. (*Rbi*,((*Igl*,*Uox*),((*Svi*,*Bta*),(*Tna*,(*Gra*,(*San*,*Gse*))))))	55,466.364	<0.001	0.114
4. (*Rbi*,((*Igl*,*Uox*),(*Bta*,(*Svi*,(*Tna*,(*San*,(*Gra*,*Gse*)))))))[c]	55,474.605	<0.001	0.047
5. (*Rbi*,((*Igl*,*Uox*),(*Bta*,(*Tna*,(*Svi*,(*Gse*,(*Gra*,*San*)))))))[d]	55,484.794	<0.001	0.010

Bta, *B. taitanus*; *Gra*, *G. ramaswamii*; *Gse*, *G. seraphini*; *Igl*, *I. glutinosus*; *Rbi*, *R. bivittatum*; *San*, *S. annulatus*; *Svi*, *S. vittatus*; *Tna*, *T. natans*; *Uox*, *U. oxyurus*.

[a]As calculated by HyPhy.

[b]Unconstrained tree (Fig. 1), Roelants et al. (2006).

[c]Wilkinson et al. (2003).

[d]Frost et al. (2006).

our data set. The recently developed non-parametric AU test uses a multiscale bootstrap approach to control for Type 1 errors, while reducing the overconservative tree selection biases of other non-parametric tests (Shimodaira, 2002). However, information about the actual power and appropriateness of this test in specific empirical cases is still limited, and several concerns exist regarding robustness to deviations from some of its basic assumptions (such as breakdown of the asymptotic theory; Shimodaira, 2002), to model misspecification (Aris-Brosou, 2003), and to heterogeneity in rates of sequence evolution (such as the effect of unequal evolutionary rates among taxa; Gissi et al., 2006).

In our case, and despite we cannot rule out some uncertainty in our caecilian tree (particularly regarding the position of *Scolecomorphus* and *Boulengerula*), the topology with *Scolecomorphus* as the sister taxon of all other higher caecilians (Fig. 1) seems to be the most robust explanation given the available data on the basis that both ML and BI methods of phylogenetic inference yield phylograms with this branch arrangement, that BI supports the arrangement with strong statistical confidence (ML only weakly), and that the PB analysis strongly rejects the alternative topologies evaluated. Besides, the full congruence of our caecilian tree with that of the recent study by Roelants et al. (in ms.) provides additional confidence on this caecilian phylogeny.

Divergence Times of Caecilian Lineages

The result of the LRT indicates non-constancy of rates of evolution among caecilian lineages (-ln $L_{\text{No Clock}}$ = 57,002.416; -ln L_{Clock} = 57,128.227; 2□ = 251.622; df = 7; $P < 0.001$). We therefore used the Bayesian relaxed clock analysis to estimate divergence times. Our results are summarized in Table 4 (node letters are cross-referenced in Figure 1) and show that the caecilian crown-group originated in the Middle-Late Jurassic (split of Rhinatrematidae 166 to 148 Mya), and that all subsequent divergences of caecilian lineages used in this study occurred during the Cretaceous (between 138 and 66 Mya). According to our estimates, the divergence of higher caecilian lineages (between 130 and 82 Mya) was concomitant with the fragmentation of Gondwana (130 to 86 Mya; Pitman III et al., 1993; Rabinowitz et al., 1983). Our results are in agreement with those of the recent study by Roelants et al. (in ms.), although the initial splits of crown-group caecilians (those of Rhinatrematidae, and the Ichthyophiidae + Uraeotyphlidae clade) are slightly older than the ones presented in this study (195-132 versus 166-132 Mya). Previous studies (San Mauro et al., 2005; Wilkinson et al., 2002) recovered even older age estimates for all caecilian divergences, suggesting that the initial crown-group split predated Pangaea fragmentation (over 180 Mya; Gurnis, 1988; Smith et al., 1994), and that even major higher caecilian splits predated Gondwana fragmentation. However, San Mauro et al.'s (2005) study also pointed out that caecilian divergences were estimated to be younger when using alternative single calibrations (means of the Rhinatrematidae split ranging from 177 to 150 Mya). The scenario inferred from San Mauro et al.'s (2005) and Wilkinson et al.'s (2002) studies allows a straightforward explanation of the current distribution of higher caecilians in South America, Africa, India, and the Seychelles, as well as of the putative African affinities of a Paleocene caeciliid-like fossil (*Apodops*) found in South America (Estes and Wake, 1972): the ancestors of major higher caecilian lineages were already distributed throughout Gondwana before its initial fragmentation (Duellman and Trueb, 1994). In contrast, our younger age estimates imply some sort of dispersion (most likely through land bridges connecting the western part of Gondwana [Africa and South America] with landmasses of the disintegrating eastern Gondwana [India, Madagascar, and the Seychelles]) to account for the current distribution of some higher caecilians, particularly Indian and Seychellean caeciliids (represented by the Indian *G. ramaswamii* in out study). Some palaeogeological models incorporating land bridges between western and eastern Gondwanan landmasses during the late Mesozoic have been recently proposed, and have helped explaining the puzzling affinities and distribution of several vertebrate groups including dinosaurs, frogs, lizards, and mammals (Briggs, 2003; Chatterjee and Scotese, 1999). Since one of the nodes constrained in our Bayesian relaxed clock analyses is indeed the one for the split between *Gegeneophis* and

TABLE 4. Divergence time estimates (in Mya) for each node of the caecilian tree (see Fig. 1).

Node	Mean	SD	95% CI
A	155.782	4.582	147.738 – 165.804
B	133.863	1.724	132.051 – 138.410
C	124.566	2.501	119.872 – 129.745
D	117.758	2.622	112.576 – 122.895
E	111.969	2.726	106.384 – 117.100
F	97.058	2.595	91.223 – 100.792
G	88.343	3.144	81.766 – 94.002
H	68.823	1.112	65.880 – 69.968

CI, credibility intervals; SD, standard deviation.

the *Geotrypetes + Siphonops* clade, we repeated the analyses removing the two time constraints on this node (leaving four constraints on two other nodes) to see the effect on resulting divergence time estimates. All re-estimated ages are virtually similar to those using time constraints on all three nodes (two to five Myr older upper bounds, same or nearly so lower bounds), with an estimated time for the split between *Gegeneophis* and the *Geotrypetes + Siphonops* clade of 106 to 92 Mya (versus 101 to 91 Mya with the full set of time constraints). Therefore, no artefacts were produced by the inclusion of age constraints on that node.

Phylogenetic Information of Mt Genes and Nuclear Rag1

The implementation of Goldman's (1998) method with Edible yielded expected phylogenetic information scores for each of the analysed data sets. These scores, along with evolutionary rates (substitutions per site, calculated as ML tree length), for each single mt and nuclear *rag1* data set are plotted in Figure 2. Both phylogenetic information and substitution rates are quite variable among the different data sets.

Substitution rate of data sets RAG1 and CO1 is relatively slower than that of all other mt data sets (Fig. 2), in agreement with previous studies that have indicated the slow evolution of nuclear *rag1* (Groth and Barrowclough, 1999; San Mauro et al., 2004b) and mt *cox1* (Lopez et al., 1997; Russo et al., 1996; San Mauro et al., 2004b; Zardoya and Meyer, 1996), this latter one particularly at amino acid level, or after exclusion of third codon positions, as in our study. A recent study (Mueller, 2006) has corroborated that *cox1*, together with the other cytochrome oxidase genes (*cox2*, *cox3*, and *cob*), possesses slow evolutionary rates at amino acid level, but also noted that they also have the fastest rates of all mt genes at nucleotide level (including all codon positions), indicating a relatively higher number of (mainly synonymous) substitutions occurring at the third position of these genes. In contrast to RAG1 or CO1, the rate of evolution of third codon positions of mt protein-coding genes (data set 3rdPOS) is extremely fast (over 100-fold faster) compared to those of all other data sets analysed (Fig. 2), which agrees with previous studies that reported the faster evolutionary rates of third codon positions with respect to first and second positions (Irwin et al., 1991; Johnson and Sorenson, 1998; Li and Graur, 1991; Rodriguez-Trelles et al., 2002). This extremely fast substitution rate is the main reason why we have separately considered all mt third codon positions (combined) as a single data set for the phylogenetic information analyses of this study.

Phylogenetic information scores reveal that the most informative single data sets for the given phylogeny are those for the tRNA genes ($2.414 \cdot 10^{17}$), *rrnS* ($1.216 \cdot 10^{16}$), *nad6* ($6.632 \cdot 10^{15}$), *rrnL* ($3.151 \cdot 10^{15}$), *nad2* ($6.405 \cdot 10^{14}$), *nad5* ($5.335 \cdot 10^{14}$), and *atp6* ($2.117 \cdot 10^{14}$) (Fig. 2). The phylogenetic performance of some of these genes, particularly ribosomal and transfer genes, is well known, and they have long been used to infer phylogenetic relationships of many diverse organisms spanning a wide range of divergence times (e.g., Cummings and Meyer, 2005; Cummings et al., 1995; Kumazawa and Nishida, 1993; Mindell and Honeycutt, 1990; Miya and Nishida, 2000; Mueller, 2006). Among the protein-coding genes, *nad2* and *nad5*, had already been indicated as good or adequate molecular markers for divergences over 300 Mya by previous studies on vertebrates (Miya and Nishida, 2000; Mueller, 2006; Russo et al., 1996; Zardoya and Meyer, 1996). In contrast, the two other genes, *nad6* and *atp6*, have usually been recovered as potentially poor (or medium at the most) phylogenetic markers (Miya and Nishida, 2000; Mueller, 2006; Zardoya and Meyer, 1996; but see Russo et al., 1996), with most studies indicating their

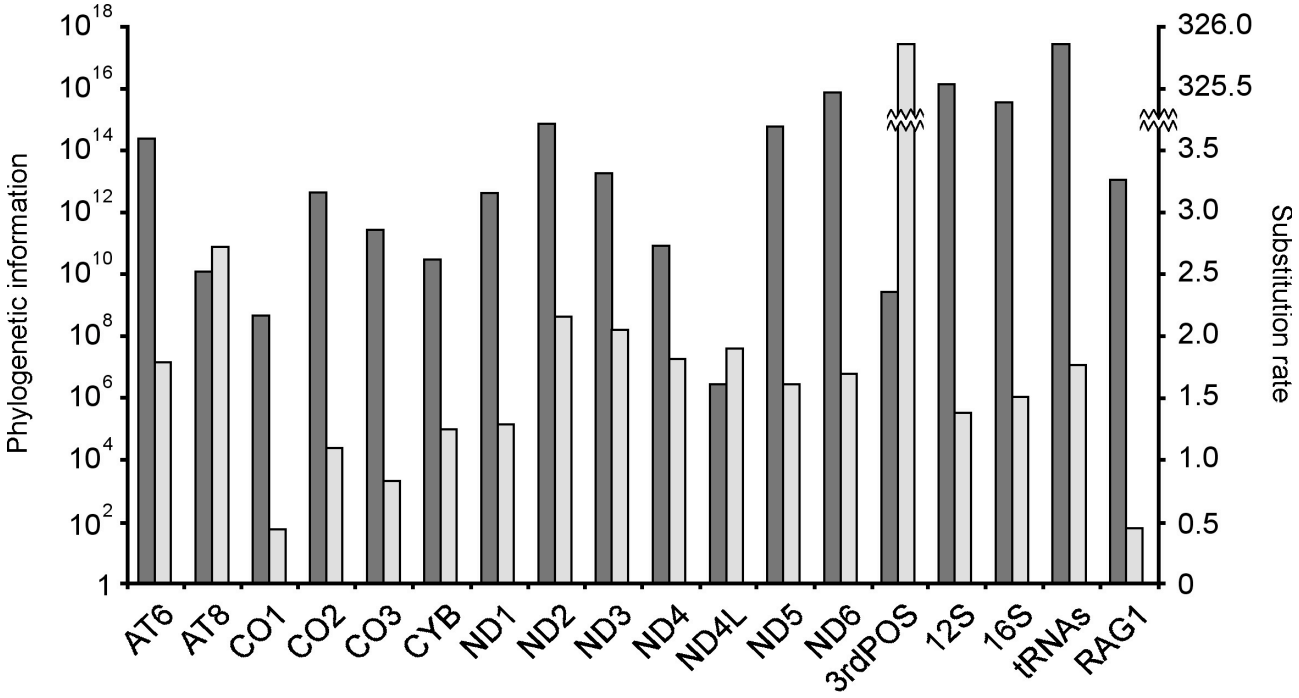


FIGURE 2. Overall phylogenetic information (dark grey bars; left) and substitution rate per site (light grey bars, right) of each single mt and nuclear *rag1* data set. Left Y-axis is on a log scale. Substitution rate is measured as ML tree length.

high variability or rate heterogeneity as probable causes eroding phylogenetic signal. Additionally, the fact that *nad6* encodes on the light strand of the mt DNA and has different base composition biases (Reyes et al., 1998) has led to this gene being routinely excluded from most phylogenetic studies using complete mt genome sequences. In contrast, our results indicate that the phylogenetic signal of this gene is relatively high (Fig. 2), and indeed comparable to that of ribosomal genes. One of the main reasons why some of our results on mt protein-coding genes are different from those of previous studies (apart from obvious differences in employed taxa) may be related with the fact that, in our study, mt protein-coding genes are examined to the exclusion of third codon positions (which are combined and analysed altogether as a single data set), thus becoming slower-evolving (with a better rate for the caecilian phylogeny at hand), and likely reducing the phylogenetic noise associated to multiple substitutions at a given position. In fact, the phylogenetic information score of third codon positions of mt protein-coding genes ($2.435 \cdot 10^9$) is among the lowest of all data sets analysed (Fig. 2), and this is probably related to their extremely fast rate of evolution (see above) that probably brings down the phylogenetic signal:noise ratio. The data set with the lowest information score is that for *nad4L* ($2.641 \cdot 10^6$), in full agreement with most previous studies (Miya and Nishida, 2000; Mueller, 2006; Russo et al., 1996; Zardoya and Meyer, 1996) that have indicated the low phylogenetic performance of this gene.

Overall Rate of Evolution and Information Scores

Phylogenetic information scores are calculated assuming an overall rate of evolution (that of the full data set comprising mt ribosomal, tRNA, and protein-coding genes, and nuclear *rag1* combined) in order to make information values comparable among the different data sets, and it is possible that this assumption may have an effect on the information calculations of data sets that notably depart from the assumed overall rate. In fact, given that the mt component of our full data set is much larger than the (slower-evolving) nuclear component (87% versus 13%), the assumed overall rate may be skewed towards the mt-component rate, therefore having an effect on the information score of the nuclear *rag1* gene (notably slower-evolving than mt data sets; Fig. 2). In order to further study the effect of the assumed overall rate of evolution on the phylogenetic information calculations, we re-estimated information scores scaling the base trees to the overall rate of the RAG1 data set (only-nuclear subset), the overall rate of the mtGENOME-NO3 data set (only-mt subset), and also the overall rate of the full data set using 16 categories to approximate the gamma-shape parameter (α_{16}). This latter re-estimation allowed us to verify whether the use of just four categories to approximate the shape parameter of the gamma distribution (as in all calculations in this study) was sufficient to yield accurate estimates of phylogenetic information.

Information scores re-estimated using overall rates

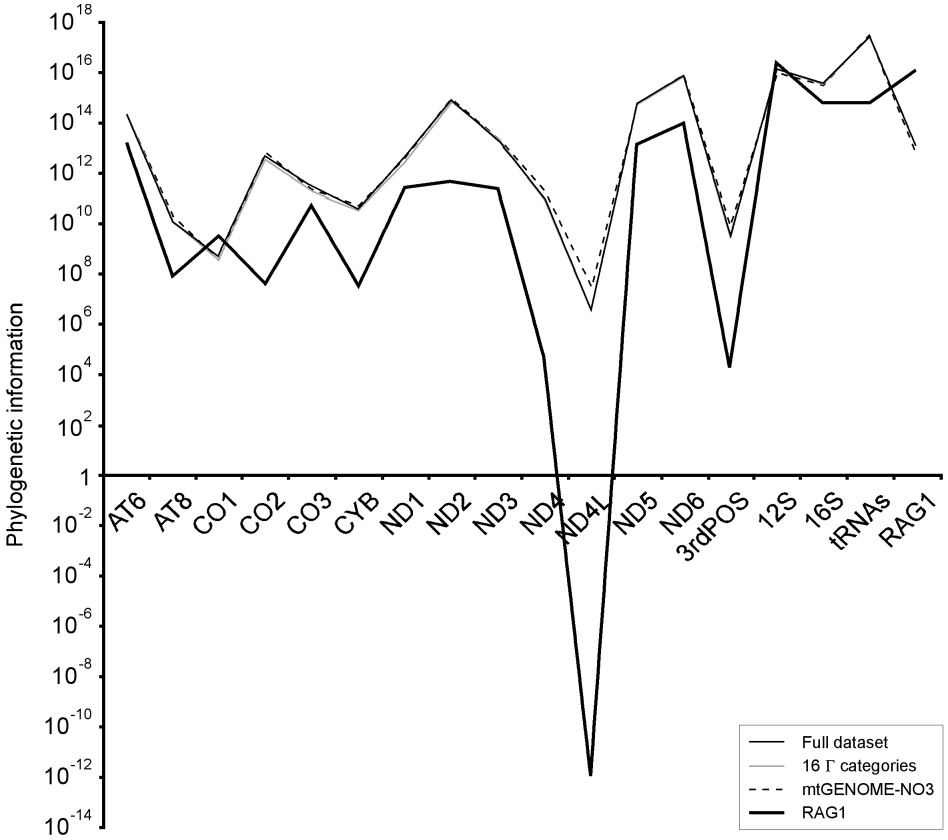


FIGURE 3. Comparison of phylogenetic information of single data sets estimated under four different tree rate optimizations (see text): that of the full data set with four Γ categories (thin black line), that of the full data set with 16 Γ categories (grey line), that of the data set including all single mt partition except that of third codon positions (dashed line), and that of the *rag1* dataset (thick black line). Y-axis is on a log scale.

of the only-mt subset or the full data set with α_{16} are virtually identical to those estimated under the original full data set (Fig. 3), with only the data set for *nad4L* doing slightly better under the only-mt-subset rate. This result indicates that four categories is adequate to approximate the gamma-shape parameter in the information calculations (at least at the scale of interest of this study), and confirms that the overall rate of the full data set is skewed towards the mt-component rate. Using overall rates of *rag1* makes a difference to the absolute information values of the different data sets (becoming lower in general since almost all data sets have a higher rate than *rag1*), but the relative pattern of values is fairly similar to that of the original full data set (Fig. 3). This similarity of pattern is indeed statistically significant with a linear regression fit of (log-transformed) information scores ($R^2 = 0.668$; $F_{1,16} = 32.153$; $P < 0.001$). In this case that the base tree is scaled to the overall rate of *rag1*, the information score of this data set is higher than that using the full data set to scale the base tree ($9.376 \cdot 10^{15}$ versus $1.020 \cdot 10^{13}$), and indeed becomes second best (after the data set for *rrnS*; Fig. 3). This is in agreement with previous studies that have indicated the high phylogenetic performance of this gene in reconstructing ancient vertebrate phylogenies (Groth and Barrowclough, 1999; Martin, 1999; San Mauro et al., 2004b; San Mauro et al., 2005). The only other notably slower-evolving data set (that of *cox1*) also has a higher absolute information score when assuming the overall rate of *rag1* ($2.570 \cdot 10^9$ versus $4.184 \cdot 10^8$; Fig. 3).

The results above suggest that assuming the overall rate of the full data set does not severely change the relative comparison of information scores among data sets, with respect to assuming the overall rate of a slow-evolving gene, such as *rag1*. Therefore, and for simplicity purposes, only the overall rate of the full data set (which, after all, encompasses rate variation of the most comprehensive set of genes) was employed in all other information analyses of this study. Since substantial differences in the relative pattern of information scores are only found in data sets with notably slower rates than the assumed overall one, i.e. *rag1* and *cox1*, some caution should be taken when comparing the overall information scores of these two particular data sets with those of all other (faster-evolving) ones.

It is worth noting the case of third codon positions of mt protein-coding genes. They are evolving so extremely fast that they should have almost no phylogenetic information whatsoever (all phylogenetic signal eroded due to saturation of substitutions). However, they actually have some information (albeit among the lowest of all data sets, but still higher than ND4L), even when scaling the base tree with the overall rate of the slow-evolving *rag1* (Figs. 2 and 3). This happens because this data set shows quite a bit of among-site rate heterogeneity (low α in the gamma distribution; Table 2), and a quarter of the sites in it (scattered throughout the 13 mt protein-coding genes) are actually evolving at about the appropriate rate (for the caecilian phylogeny at hand), with the other three quarters of the sites going so extremely fast that do not virtually add anything to the overall information score. However, the fact that the vast majority of sites do not add information (because the "true" phylogenetic signal is mostly lost in them) does not necessarily prevent them from generating false, spurious signals (due to accumulation of superimposed, homoplasic nucleotide substitutions) that confound phylogenetic inference, and yield artefactual reconstructions. In fact, the inclusion of third codon positions in our phylogenetic analyses yields artefactual topology reconstructions (e.g., Typhlonectidae recovered at the base of the caecilian tree, branching off even before the Ichthyophiidae + Uraeotyphlidae clade) and/or reduces the statistical support for otherwise well-supported relationships, irrespective of the method of phylogenetic inference employed (not shown).

Phylogenetic Information Per Branch

Apart from the overall quantification, information scores were also estimated per branch of the unrooted caecilian tree, allowing us to compare expected phylogenetic information among data sets in specific parts of the tree. Figure 4 shows these per-branch information scores for each single data set. Instead of an unrooted tree, the tree shown in Figure 4 is the ultrametric tree resulting from the Bayesian relaxed clock analysis, hence allowing direct confrontation of information scores with branch position as well as with branch age.

In general, information scores of all data sets are lower in terminal than in the most internal branches, those spanning a time depth of 138 to 82 Myr. The branches gathering the highest information scores are those internals for the splits of *Boulengerula* and *Typhlonectes* (covering a time spam of 123 to 106 Mya). As for the overall quantification, the most informative data sets in most branches are those for the tRNA and ribosomal genes, and those for *nad6*, *nad5*, *nad2*, and *atp6*. Although there seems to be a general pattern of relative data set information across branches, it changes slightly in each branch, with some data sets performing better in some branches, but becoming worse in some others (Fig. 4). Interestingly, the relative performance of slow-evolving (those for *cox1* and *rag1*) and fast-evolving (that of third codon positions) data sets changes more markedly between internal and terminal branches. We conducted a factorial (two-way) analysis of variance to assess variations in (log-transformed) phylogenetic information between terminal and internal branches (main effect "Branch Type"), and between slow-evolving (those for *cox1* and *rag1*), fast-evolving (that of third codon positions), and all other data sets (main effect "Gene Rate"). Both main effects are highly significant ($F_{1,264} = 128.619$ for "Branch Type"; $F_{1,264} = 18.187$ for "Gene Rate"; $P < 0.001$ in both cases), indicating that information scores are significantly higher in internal than in terminal branches, and that, in general, fast- and slow-evolving data sets perform worse than other data sets. However, the interaction of the two main effects is also significant ($F_{1,264} = 3.348$; $P = 0.037$), and we used planned comparisons to examine contrasts between slow-evolving and other data sets in both internal and in terminal branches, and between fast-evolving and other data sets also in both internal and in terminal branches. Differences between slow-evolving and other data sets were significant in terminal branches ($F_{1,264} = 6.085$; $P = 0.014$), but non-significant in internal branches ($F_{1,264} = 1.759$; $P = 0.186$). In contrast, differences between fast-evolving and other data sets were non-significant in terminal branches ($F_{1,264} = 1.009$; $P = 0.316$), but significant in internal branches ($F_{1,264} =$

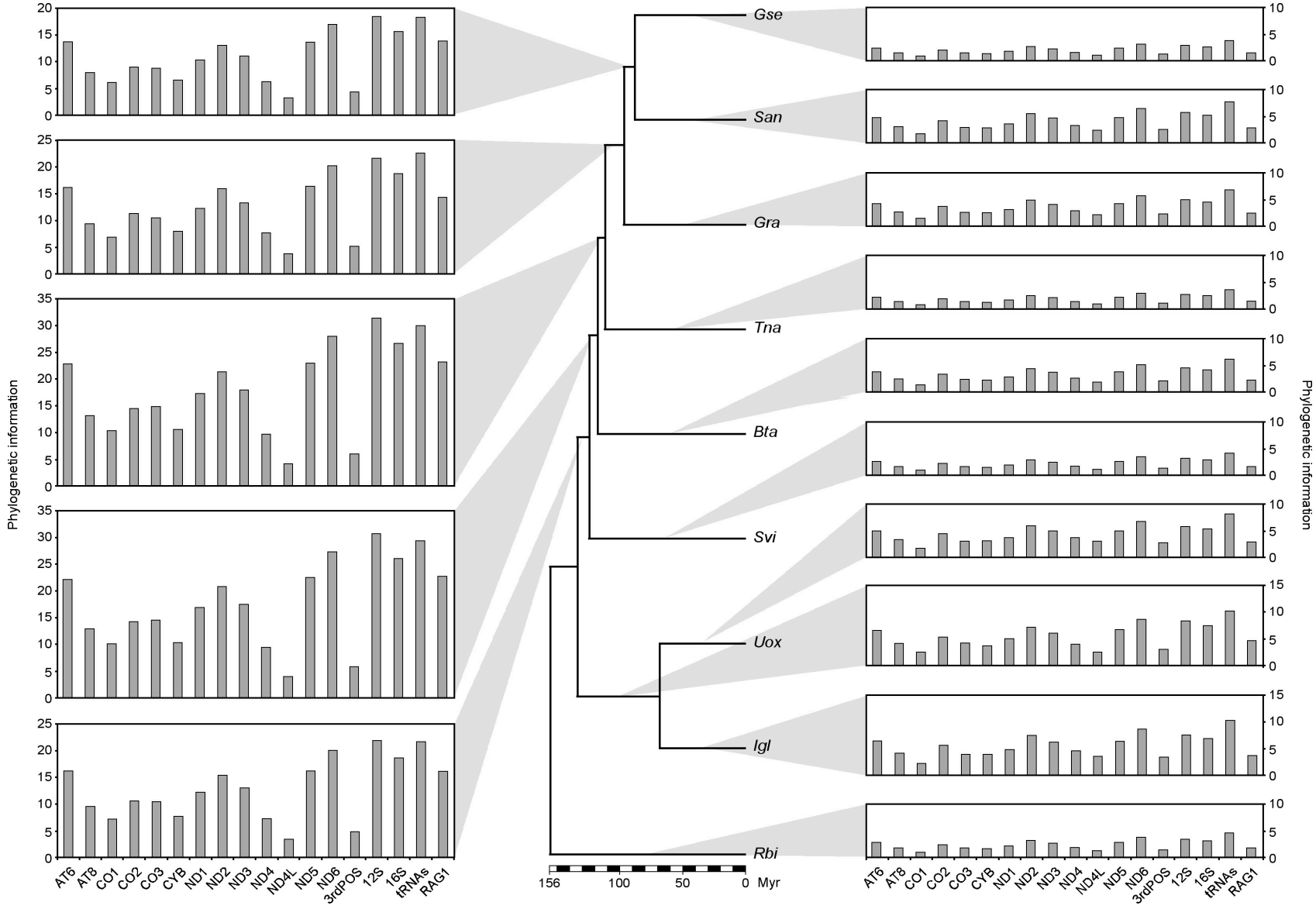


FIGURE 4. Phylogenetic information content of single data sets estimated per branch of the caecilian tree. The tree depicts divergence time estimates (in millions of years, Myr) for each node (see also Table 4). *Bta*, *B. taitanus*; *Gra*, *G. ramaswamii*; *Gse*, *G. seraphini*; *Igl*, *I. glutinosus*; *Rbi*, *R. bivittatum*; *San*, *S. annulatus*; *Svi*, *S. vittatus*; *Tna*, *T. natans*; *Uox*, *U. oxyurus*.

11.949; $P < 0.001$). This indicates that data sets for *cox1* and *rag1* perform as good or bad as all other faster-evolving data sets in the internal branches, but that they perform significantly worse than faster-evolving data sets in the terminal branches of the tree. Conversely, fast-evolving third codon positions perform relatively worse than other slower-evolving data sets in the internal branches of the tree, even though they have higher absolute information scores in internal than in terminal branches (Fig. 4).

Combining Information of Mt Data Sets: Assessing Mitogenomic Information

As for single data sets, phylogenetic information can be also quantified for joint data sets involving combination of single smaller data sets. In this sense, the combination of all single mt data sets (for ribosomal, tRNA, and protein-coding genes) yields an estimate of the overall phylogenetic information of the complete mt genome. However, assessment of phylogenetic information of combined data sets cannot be done in every way. If single data sets are first concatenated into a combined one, and then information is estimated, the consequent averaging of the substitution model parameters in the joint data set may mislead the quantification of information. Rather, phylogenetic information of a combined data set should be estimated as the addition of the information scores of each single data set in the combined one.

In order to explore the effect of model averaging, phylogenetic information of combined mt data sets was estimated directly from concatenated data sets (PROTS-NO3, PROTS-ALL, and mtGENOME-NO3; Table 2), as well as adding up the information scores of contributing single data sets. The results show that there is a notable variation in information scores between those data sets averaging phylogenetic information, and those adding up information (Fig. 5). For example, data sets PROTS-ALL and PROTS-NO3 + 3rdPOS are based on the same set of sequence characters, but information scores are markedly different ($3.566 \cdot 10^{15}$ versus $3.675 \cdot 10^{14}$). Also, the sum of information scores of all single mt protein-coding data sets (including third positions) yields a different overall score ($8.042 \cdot 10^{15}$), even though the set of sequence characters is the same as above. Similarly, phylogenetic information of data set mtGENOME-NO3 is largely different form that calculated adding up information scores of all single mt data sets (excluding third positions), despite they both consist on the same mitogenomic set of sequence characters ($5.809 \cdot 10^{15}$ versus $2.647 \cdot 10^{17}$). In fact, information score of data set mtGENOME-NO3 is lower than that of e.g. the data set for *rrnS* ($1.216 \cdot 10^{16}$), which is rather illogical given that *rrnS* is part of data set mtGENOME-NO3.

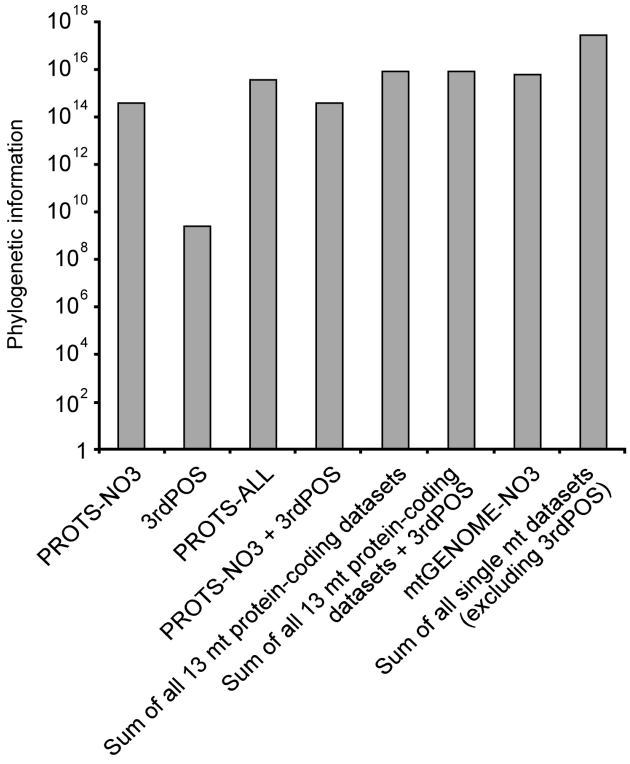


FIGURE 5. Phylogenetic information scores of data sets involving combination of smaller single mt data sets. Y-axis is on a log scale.

In general, and apart from the quantification of overall mitogenomic information mentioned above, information scores of every single data set can be combined with to those of any other (adding up, and not averaging) to get joint estimates of phylogenetic information. For example, this can be used to quantify information of genes with functionally related peptide products (such as the subunits of a protein complex), or genes that are located in specific regions of the mt genome.

Informative Data Sets and the Uncertain Position of Scolecomorphus and Boulengerula

In general, the most informative genes for the employed caecilian phylogeny, spanning time divergences of 166 to 66 Mya, are those for tRNA and ribosomal genes. Also, some genes encoding subunits of the NADH dehydrogenase complex (*nad2*, *nad5*, and interestingly *nad6*) and the nuclear slower-evolving *rag1* are very informative, this latter one particularly in the more internal branches of the tree. An experimental design including the most informative data set evaluated, that for the mt tRNA genes, which account for a large part of the mitogenomic information, would perhaps be optimal. However, the reduced size of individual tRNA genes, along with the fact that they are scattered all over the caecilian mt genome (San Mauro et al., 2004b; Zardoya and Meyer, 2000) makes this data set difficult to be obtained separately from the rest of the mt genome (at least using standard PCR-based approaches for determining sequences, as most molecular phylogenetic studies do).

Obtaining whole mitogenomic data makes available additional information (other than just that of sequence characters), such as gene order, that in some cases (when rearranged) may provide evidence of phylogenetic relationships, and may help to better understand the mechanisms driving the evolution of the mt genome (Rokas and Holland, 2000; San Mauro et al., 2006). However, the mitogenomic option may not always be feasible (or even necessary) so, given the location of those most-informative mt genes in the mt genome, a reasonable design to study overall caecilian phylogenetic questions might be the sequencing of the mt region covering both complete ribosomal genes, the mt region covering *nad5* and *nad6* (both genes found next to each other in all caecilian mt genomes so far; San Mauro et al., 2004b; Zardoya and Meyer, 2000), and

perhaps the nuclear *rag1* fragment (which, in addition, adds information of different genomic source) particularly if we are interested in the deeper internal branches.

Interestingly, phylogenetic reconstructions (ML and BI with same settings as indicated for the full data set) based on this "informative" data set (*rrnS*, *rrnL*, *nad5*, *nad6*, *rag1*) yielded a strongly supported tree (all node support values being maximal or nearly so for BI and high [>80] to maximal for ML) that is similar in topology to that of Fig. 1, except for *Scolecomorphus* and *Boulengerula* that have swap over positions, so that *Boulengerula* appears as the sister taxon of all other caecilians, and *Scolecomorphus* branches off the tree next (topology 2 in Table 3). As mentioned above, most of the uncertainty of our caecilian phylogeny (both from ML phylogenetic analyses and AU tests of alternative topologies) is on this node, and the position of *Scolecomorphus* and *Boulengerula* appears also highly controversial in the literature (Frost et al., 2006; Roelants et al., in ms.; Wilkinson et al., 2003).

Taxon Sampling Strategy

We used Goldman's (1998) method to identify branches in the caecilian tree where is best to add taxa in order to increase phylogenetic information of the controversial *Boulengerula-Scolecomorphus* node. Figure 6 shows the variation in phylogenetic information for the branch leading to this node when adding new hypothetical taxa in different branches of the tree, and at different positions along each branch (measured as the distance along the edge the adding has happened). For all branches, the gain of information is strongly correlated with the distance ($R^2 > 0.980$; $F_{1,10} > 489.634$; $P < 0.001$ in all cases) with information typically going up as the additional taxa joins the tree nearer the controversial node. We conducted an analysis of covariance (distance as covariate) to assess variations in (log-transformed) phylogenetic information, and used planned comparisons to examine contrasts between adding taxa at specific branches.

The greatest gain in information (significantly higher than those in all other branches; $F_{1,153} = 639.285$; $P < 0.001$) occurs when the hypothetical taxon joins internal branch 1 neighbouring the controversial node (phylogenetic information going up to 28.395) (Fig. 6), in agreement with Goldman (1998) who indicated that the nearer to the node (of interest) that the additional taxa joins the tree, the greater the gain in information (becoming maximal when the new taxa joins exactly at the node). Unfortunately, in terms of known caecilian diversity, and despite the fact that the caecilian phylogeny is not fully resolved yet, it seems unlikely that a caecilian taxon joining the tree at this (or, in general, at any other) internal branch can be found (Wilkinson and Nussbaum, 2006). Most likely, the caecilian taxa known so far will fit in the tree as sisters of the terminal branches. Of these, significant increases in information ($F_{1,153} = 172.809$; $P < 0.001$) occur when the hypothetical taxon is added to the branch of *Scolecomorphus* (going up to 28.483) and the branch of *Boulengerula* (going up to 26.815), likely due to the bisection of the terminal branches directly leading to the controversial node, and interestingly to the branch of *Rhinatrema* (going up to 26.182), likely due to the stabilization of the root of the caecilian tree (Fig. 6). When the hypothetical taxa is added to any other terminal branch, the increase in information is not significant, and the values stay around the score without adding any additional taxa (23.442).

All these results are in agreement with previous studies (Graybeal, 1998; Hillis, 1998; Poe and Swofford, 1999; Pollock and Bruno, 2000; Pollock et al., 2002; Rannala et al., 1998; Zwickl and Hillis, 2002) that suggested that increasing taxon sampling improves overall phylogenetic accuracy (but see Kim, 1998; Rokas and Carroll, 2005; Rosenberg and Kumar, 2001). Perhaps more importantly, these results provide clues on where in the caecilian tree is best to add new taxa

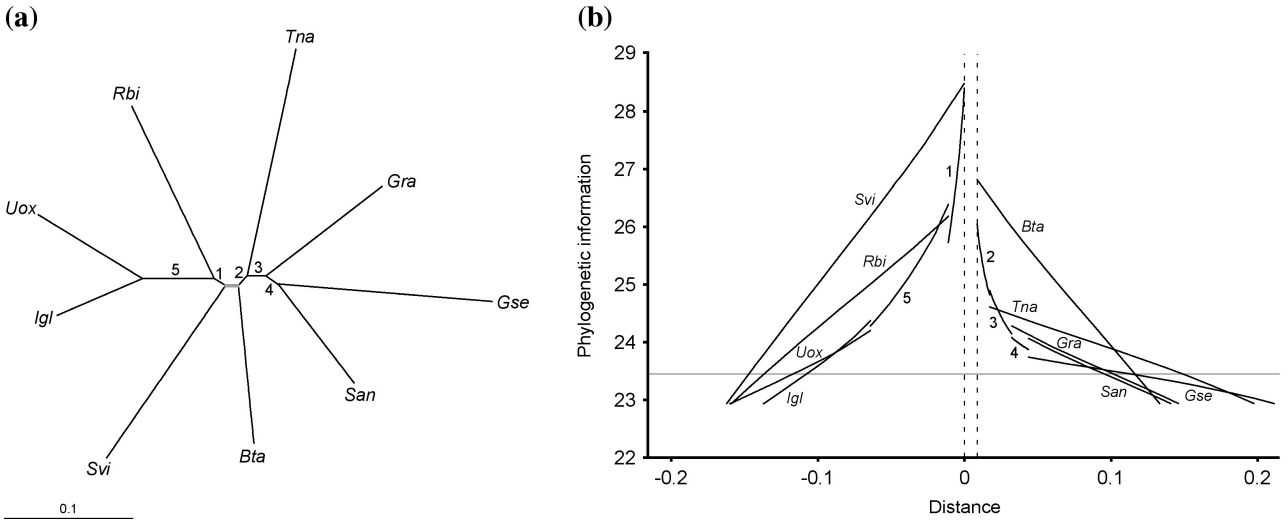


FIGURE 6. Variation in phylogenetic information for the controversial *Boulengerula-Scolecomorphus* branch as new hypothetical taxa are added to the caecilian tree. (a) The model caecilian phylogeny, indicating the controversial branch (thick grey line) as described in the text, and with branches labelled. (b) Phylogenetic information plotted against the distance along the branch the adding has happened. Regions of the graph are labelled according to which branch on the model caecilian tree the new taxon is attached to. Vertical dashed lines denote the boundaries of the controversial branch. Lower horizontal grey line indicates the information content of the controversial branch without adding any additional taxa. *Bta*, *B. taitanus*; *Gra*, *G. ramaswamii*; *Gse*, *G. seraphini*; *Igl*, *I. glutinosus*; *Rbi*, *R. bivittatum*; *San*, *S. annulatus*; *Svi*, *S. vittatus*; *Tna*, *T. natans*; *Uox*, *U. oxyurus*.

for tackling the controversial phylogenetic position of *Scolecomorphus* and *Boulengerula*, and the underlying paraphyly of Caeciliidae with respect to Scolecomorphidae. According to our results, key caecilian genera that would potentially increase phylogenetic accuracy in this part of the caecilian tree, and that should certainly be targeted in future phylogenetic studies are the sister taxa of *Scolecomorphus*, *Boulengerula*, and *Rhinatrema*, which, according to recent studies (Frost et al., 2006; Roelants et al., in ms.; Wilkinson and Nussbaum, 2006), appear to be *Crotaphatrema*, *Herpele*, and *Epicrionops*, respectively.

Experimental Design and Caecilian Systematics

The information presented in this study allows the assessment of the expected phylogenetic information that a particular combination of genes might yield, and the taxa that should be next targeted to stabilize the controversial *Boulengerula-Scolecomorphus* node. It would therefore allow an *a priori* evaluation of the appropriateness of particular experimental designs to solve specific questions at different levels of the caecilian phylogeny. Moreover, and given that mt molecular evolutionary patterns are often quite similar across vertebrates (Wolstenholme, 1992), the results of our study (or at least some of them) might perhaps be extended (always with caution) to other groups with similar divergence time spans and rates of molecular evolution.

One should keep in mind, however, that Goldman's (1998) method only allows quantification of expected phylogenetic information for a set of particular model parameters and base tree. Thus, even the use of the data sets and taxon sampling with the highest scores here presented does not guarantee a fully resolved phylogeny in our future study, but rather it ensures to have quantitatively selected the genes or taxon sampling combination that *a priori* maximize experimental design, and thus may increase the chances that our future study will be a successful one.

ACKNOWLEDGEMENTS

We are grateful to Tim Massingham for invaluable mathematical and programming assistance with the calculations of phylogenetic information scores. This work received financial support from grants of the Ministry of Education and Science of Spain (CGL2004-00401, and FPI PhD fellowship BES-2002-2033), the Natural Environment Research Council (GST/02/832) and the Biotechnology and Biological Sciences Research Council (40/G18385) of the United Kingdom, the Science Foundation Ireland (EEB0026), and the European Commission's Research Infrastructure Action via the SYNTHESYS Project.

REFERENCES

Akaike, H. 1973. Information theory as an extension of the maximum likelihood principle*in* Second international symposium of information theory (B. N. Petrov, and F. Csaki, eds.). Akademiai Kiado, Budapest, Hungary.

Altschul, S. F., W. Gish, W. Miller, E. W. Myers, and D. J. Lipman. 1990. Basic local alignment search tool. J. Mol. Biol. 215:403-410.

AmphibiaWeb. 2006. AmphibiaWeb: Information on amphibian biology and conservation. [web application] (Accessed: Sep 18, 2006) http://amphibiaweb.org/ Berkeley, CA.

Aris-Brosou, S. 2003. Least and most powerful tests to elucidate the origin of seed plants in the presence of conflicting signals under misspecified models. Syst. Biol. 52:781-793.

Atkinson, A. C., and A. N. Donev. 1992. Optimum experimental designs. Oxford University Press, London.

Bonferroni, C. E. 1936. Teoria statistica delle classi e calcolo delle probabilità. Pubblicazioni del R Istituto Superiore di Scienze Economiche e Commerciali di Firenze 8:3-62.

Brennicke, A., and D. A. Clayton. 1981. Nucleotide assignment of alkalisensitive sites in mouse mitochondrial DNA. J. Biol. Chem. 256:10613–10617.

Briggs, J. C. 2003. The biogeographic and tectonic history of India. J. Biogeogr. 30:381-388.

Buckley, T. R. 2002. Model misspecification and probabilistic tests of topology: evidence from empirical data sets. Syst. Biol. 51:509-523.

Castresana, J. 2000. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. Mol. Biol. Evol. 17:540-552.

Chatterjee, S., and C. R. Scotese. 1999. The breakup of Gondwana and the evolution and biogeography of the Indian plate. Proc. Indian Natl. Sci. Acad. 65A:397-425.

Clary, D. O., and D. R. Wolstenholme. 1985. The mitochondrial DNA molecule of *Drosophila yakuba*: nucleotide sequence, gene organization and genetic code. J. Mol. Evol. 22:252-271.

Clayton, D. A. 1982. Replication of animal mitochondrial DNA. Cell 28:693-705.

Cummings, M. P., and A. Meyer. 2005. Magic bullets and golden rules: data sampling in molecular phylogenetics. Zoology 108:329-336.

Cummings, M. P., S. P. Otto, and J. Wakeley. 1995. Sampling properties of DNA sequence data in phylogenetic analysis. Mol. Biol. Evol. 12:814-822.

Curole, J. P., and T. D. Kocher. 1999. Mitogenomics: digging deeper with complete mitochondrial genomes. Trends Ecol. Evol. 14:394-398.

Duellman, W. E., and L. Trueb. 1994. Biology of amphibians. Johns Hopkins University Press, Baltimore, MD.

Edwards, A. W. F. 1972. Likelihood. Cambridge University Press, Cambridge.

Efron, B. 1985. Bootstrap confidence intervals for a class of parametric problems. Biometrika 72:45-58.

Estes, R., and M. H. Wake. 1972. The first fossil record of caecilian amphibians. Nature 239:228-231.

Felsenstein, J. 1981. Evolutionary trees from DNA sequences: a maximum likelihood approach. J. Mol. Evol. 17:368-376.

Frost, D. R. 2004. Amphibian species of the World: an online reference. V3.0 (22 August, 2004). http://research.amnh.org/herpetology/amphibia/index.html American Museum of Natural History, New York, NY.

Frost, D. R., T. Grant, J. Faivovich, R. H. Bain, A. Haas, C. F. B. Haddad, R. O. de Sá, A. Channing, M. Wilkinson, S. C. Donnellan, C. J. Raxworthy, J. A. Campbell, B. L. Blotto, P. Moler, R. C. Drewes, R. A. Nussbaum, J. D. Lynch, D. M. Green, and W. C. Wheeler. 2006. The amphibian tree of life. Bull. Am. Mus. Nat. His. 297:1-370.

Gissi, C., D. San Mauro, G. Pesole, and R. Zardoya. 2006. Mitochondrial phylogeny of Anura (Amphibia): A case study of congruent phylogenetic reconstruction using amino acid and nucleotide characters. Gene 366:228-237.

Goldman, N. 1993. Statistical tests of models of DNA substitution. J. Mol. Evol. 36:182–198.

Goldman, N. 1998. Phylogenetic information and experimental design in molecular systematics. Proc. R. Soc. Lond. B 265:1779-1786.

Goldman, N., J. P. Anderson, and A. G. Rodrigo. 2000. Likelihood-based tests of topologies in phylogenetics. Syst. Biol. 49:652-670.

Gower, D. J., S. P. Loader, C. B. Moncrieff, and M. Wilkinson. 2004. Niche separation and comparative abundance of *Boulengerula boulengeri* and *Scolecomorphus vittatus* (Amphibia: Gymnophiona) in an East Usambara forest, Tanzania. Afr. J. Herpetol. 53:183-190.

Graybeal, A. 1994. Evaluating the phylogenetic utility of genes: a search for genes informative about deep divergences among vertebrates. Syst. Biol. 43:174-193.

Graybeal, A. 1998. Is it better to add taxa or characters to a difficult phylogenetic problem? Syst. Biol. 47:9-17.

Groth, J. G., and G. F. Barrowclough. 1999. Basal divergences in birds and the phylogenetic utility of the nuclear RAG-1 gene. Mol. Phylogenet. Evol. 12:115-123.

Gurnis, M. 1988. Large-scale mantle convection and the aggregation and dispersal of supercontinents. Nature 322:695–699.

Hedges, S. B., R. A. Nussbaum, and L. R. Maxson. 1993. Caecilian phylogeny and biogeography inferred from mitochondrial DNA sequences of the 12SrRNA and 16S rRNA genes (Amphibia: Gymnophiona). Herpetol. Monogr. 7:64-76.

Hillis, D. M. 1998. Taxonomic sampling, phylogenetic accuracy, and investigatior bias. Syst. Biol. 47:3-8.

Hixson, J. E., T. W. Wong, and D. A. Clayton. 1986. Both the conserved stem-loop and divergent 5'-flanking sequences are required for initiation at the human mitochondrial origin of light-strand DNA replication. J. Biol. Chem. 261:2384-2390.

Huelsenbeck, J. P., and K. A. Crandall. 1997. Phylogeny estimation and hypothesis testing using maximum likelihood. Ann. Rev.

Ecol. Syst. 28:437-466.
Huelsenbeck, J. P., D. M. Hillis, and R. Jones. 1996a. Parametric bootstrapping in molecular phylogenetics: Applications and performance. Pages 19-45 *in* Molecular Zoology: Advances, Strategies, and Protocols (J. D. Ferarris, and S. R. Palumbi, eds.). Wiley-Liss, New York.
Huelsenbeck, J. P., D. M. Hillis, and R. Nielsen. 1996b. A likelihood ratio test of monophyly. Syst. Biol. 45:546-558.
Huelsenbeck, J. P., and F. R. Ronquist. 2001. MRBAYES: Bayesian inference of phylogenetic trees. Bioinformatics 17:754-755.
Huelsenbeck, J. P., F. R. Ronquist, R. Nielsen, and J. P. Bollback. 2001. Bayesian inference of phylogeny and its impact on evolutionary biology. Science 294:2310-2314.
Irwin, D. M., T. D. Kocher, and A. C. Wilson. 1991. Evolution of the cytochrome *b* gene of mammals. J. Mol. Evol. 32:128-144.
Jameson, D., A. P. Gibson, C. Hudelot, and P. G. Higgs. 2003. OGRe: a relational database for comparative analyses of mitochondrial genomes. Nucleic Acids Res. 31:202-206.
Johnson, K. P., and M. D. Sorenson. 1998. Comparing Molecular Evolution in Two Mitochondrial Protein Coding Genes (Cytochrome b and ND2) in the Dabbling Ducks (Tribe: Anatini). Mol. Phylogenet. Evol. 10:82-94.
Kim, J. 1996. General inconsistency conditions for maximum parsimony: effects of branch lengths and incresing numbers of taxa. Syst. Biol. 45:363-374.
Kim, J. 1998. Large-scale phylogenies and measuring the performance of phylogenetic estimators. Syst. Biol. 47:43-60.
Kishino, H., J. L. Thorne, and W. J. Bruno. 2001. Performance of a divergence time estimation method under a probabilistic model of rate evolution. Mol. Biol. Evol. 18:352-361.
Kosakovsky Pond, S. L., S. D. W. Frost, and S. V. Muse. 2005. HyPhy: hypothesis testing using phylogenies. Bioinformatics 21:676-679.
Kumazawa, Y. 2004. Mitochondrial DNA sequences of five squamates: phylogenetic affiliation of snakes. DNA Res. 11:137-144.
Kumazawa, Y., and M. Nishida. 1993. Sequence evolution of mitochondrial tRNA genes and deep-branch animal phylogenetics. J. Mol. Evol. 37:380-398.
Kumazawa, Y., H. Ota, M. Nishida, and T. Ozawa. 1996. Gene rearrangements in snake mitochondrial genomes: highly concerted evolution of control-region-like sequences duplicated and inserted into a tRNA cluster. Mol. Biol. Evol. 13:1242-1254.
Kumazawa, Y., H. Ota, M. Nishida, and T. Ozawa. 1998. The complete nucleotide sequence of snake (*Dinodon semicarinatus*) mitochondrial genome with two identical control regions. Genetics 150:313-329.
Kupfer, A., H. Müller, M. M. Antoniazzi, C. Jared, H. Greven, R. A. Nussbaum, and M. Wilkinson. 2006. Parental investment by skin feeding in a caecilian amphibian. Nature 440:926-929.
Li, W.-H., and D. Graur. 1991. Fundamentals of Molecular Evolution. Sinauer, Sunderland, MA.
Loader, S. P., M. Wilkinson, D. J. Gower, and C. A. Msuya. 2003. A remarkable young *Scolecomorphus vittatus* (Amphibia: Gymnophiona: Scolecomorphidae) from the North Pare Mountains, Tanzania. J. Zool. 259:93-101.
Lopez, J. V., M. Culver, J. C. Stephens, W. E. Johnson, and S. J. O'Brien. 1997. Rates of nuclear and cytoplasmic mitochondrial DNA sequence divergence in mammals. Mol. Biol. Evol. 14:277-286.
Lunt, D. H., and B. C. Hyman. 1997. Animal mitochondrial DNA recombination. Nature 387:247.
Macey, J. R., A. Larson, N. B. Ananjeva, Z. Fang, and T. J. Papenfuss. 1997. Two novel gene orders and the role of light-strand replication in rearrengement of the vertebrate mitochondrial genome. Mol. Biol. Evol. 14:91-104.
Macey, J. R., J. A. Schulte II, A. Larson, and T. J. Papenfuss. 1998. Tandem duplication via light-strand synthesis may provide a precursor for mitochondrial genomic rearrangement. Mol. Biol. Evol. 15:71-75.
Martin, A. P. 1999. Susbtitution rates of organelle and nuclear genes in sharks: implicating metabolic rate (again). Mol. Biol. Evol. 16:996-1002.
Massingham, T., and N. Goldman. 2000. EDIBLE: experimental design and information calculations in phylogenetics. Bioinformatics 16:294-295.
Mindell, D. P., and R. L. Honeycutt. 1990. Ribosomal RNA in vertebrates: evolution and phylogenetic applications. Ann. Rev. Ecol. Syst. 21:541-566.
Mindell, D. P., M. D. Sorenson, and D. E. Dimcheff. 1998. Multiple independent origins of mitochondrial gene order in birds. Proc. Natl. Acad. Sci. USA 95:10693-10697.
Miya, M., and M. Nishida. 2000. Use of mitogenomic information in teleostean molecular phylogenetics: a tree-based exploration under the maximum-parsimony optimality criterion. Mol. Phylogenet. Evol. 17:437-455.
Moritz, C., and W. M. Brown. 1987. Tandem duplications in animal mitochondrial DNAs: variation in incidence and gene content among lizards. Proc. Natl. Acad. Sci. USA 84:7183-7187.
Moritz, C., T. E. Dowling, and W. M. Brown. 1987. Evolution of animal mitochondrial DNA: relevance for population biology and systematics. Annu. Rev. Ecol. Syst. 18:269-292.
Mueller, R. L. 2006. Evolutionary rates, divergence dates, and the performance of mitochondrial genes in Bayesian phylogenetic analysis. Syst. Biol. 55:289-300.
Müller, H. 2006. Ontogeny of the skull, lower jaw, and hyobranchial skeleton of *Hypogeophis rostratus* (Amphibia: Gymnophiona: Caeciliidae) revisited. J. Morph. 267:968-986.
Müller, H., O. V. Oommen, and P. Bartsch. 2005. Skeletal development of the direct developing caecilian *Gegeneophis ramaswamii* (Amphibia: Gymnophiona: Caeciliidae). Zoomorphology 124:171—188.
Nussbaum, R. A. 1977. Rhinatrematidae: a new family of caecilians (Amphibia: Gymnophiona). Occ. Pap. Mus. Zool. Univ. Michigan 682:1-30.
Nussbaum, R. A. 1979. The taxonomic status of the caecilian genus *Uraeotyphlus* Peters. Occ. Pap. Mus. Zool. Univ. Michigan 687:1-20.
Nussbaum, R. A., and M. Wilkinson. 1989. On the classification and phylogeny of caecilians (Amphibia: Gymnophiona), a critical review. Herpetol. Monogr. 3:1-42.
Pitman III, W. C., S. Cande, J. LaBrecque, and J. Pindell. 1993. Fragmentation of Gondwana: the separation of Africa from South America. Pages 15-34 *in* Biological relationships between Africa and South America (P. Goldblatt, ed.) Yale University Press, New Haven, CT.
Poe, S., and D. L. Swofford. 1999. Taxon sampling revisited. Nature 398:299-300.
Pollock, D. D., and W. J. Bruno. 2000. Assessing an unknown evolutionary process: effect of increasing site-specific knowledge through taxon addition. Mol. Biol. Evol. 17:1854-1858.
Pollock, D. D., D. J. Zwickl, J. A. McGuire, and D. M. Hillis. 2002. Increased taxon sampling is advantageous for phylogenetic inference. Syst. Biol. 51:664-671.
Posada, D., and K. A. Crandall. 1998. MODELTEST: testing the model of DNA substitution. Bioinformatics 14:817-818.
Rabinowitz, P. D., M. F. Coffin, and D. Falvey. 1983. The separation of Madagascar and Africa. Science 220:67-69.
Rannala, B., J. P. Huelsenbeck, Z. Yang, and R. Nielsen. 1998. Taxon sampling and the accuracy of large phylogenies. Syst. Biol. 47:702-710.
Reeves, J. H. 1992. Heterogeneity in the substitution process of amino acid sites of proteins coded for by mitochondrial DNA. J. Mol. Evol. 35:17-31.
Reyes, A., C. Gissi, G. Pesole, and C. Saccone. 1998. Asymetrical directional mutation pressure in the mitochondrial genome of mammals. Mol. Biol. Evol. 15:957-966.
Rodriguez, F., J. F. Oliver, A. Marín, and J. R. Medina. 1990. The general stochastic model of nucleotide substitution. J. Theor. Biol. 142:485-501.
Rodriguez-Trelles, F., L. Alarcón, and A. Fontdevila. 2002. Molecular Evolution and Phylogeny of the *buzzatii* Complex (*Drosophila repleta* Group): A Maximum-Likelihood Approach. Mol. Biol. Evol. 17:1112-1122.
Roelants, K., D. J. Gower, M. Wilkinson, S. P. Loader, S. D. Biju, K. Guillaume, and F. Bossuyt. High extinction rates and fluctuating diversification in the history of modern amphibians. Unpublished manuscript.
Rokas, A., and S. B. Carroll. 2005. More genes or more taxa? The relative contribution of gene number and taxon number to phylogenetic accuracy. Mol. Biol. Evol. 22:1337-1344.
Rokas, A., and P. W. H. Holland. 2000. Rare genomic changes as a tool for phylogenetics. Trends Ecol. Evol. 15:454-459.
Ronquist, F., and J. P. Huelsenbeck. 2003. MRBAYES 3: Bayesian phylogenetic inference under mixed models. Bioinformatics 19:1572-1574.
Rosenberg, M. S., and S. Kumar. 2001. Incomplete taxon sampling is not a problem for phylogenetic inference. Proc. Natl. Acad. Sci. USA 98:10751-10756.
Russo, C. A. M., N. Takezaki, and M. Nei. 1996. Efficiencies of different genes and different tree-building methods in recovering a known vertebrate phylogeny. Mol. Biol. Evol. 13:525-536.
Sambrook, J., E. F. Fritsch, and T. Maniatis. 1989. Molecular cloning. A laboratory manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
San Mauro, D., M. Garcia-Paris, and R. Zardoya. 2004a. Phylogenetic relationships of discoglossid frogs (Amphibia: Anura: Discoglossidae) based on complete mitochondrial genomes and nuclear genes. Gene 343:357-366.
San Mauro, D., D. J. Gower, O. V. Oommen, M. Wilkinson, and R. Zardoya. 2004b. Phylogeny of caecilian amphibians

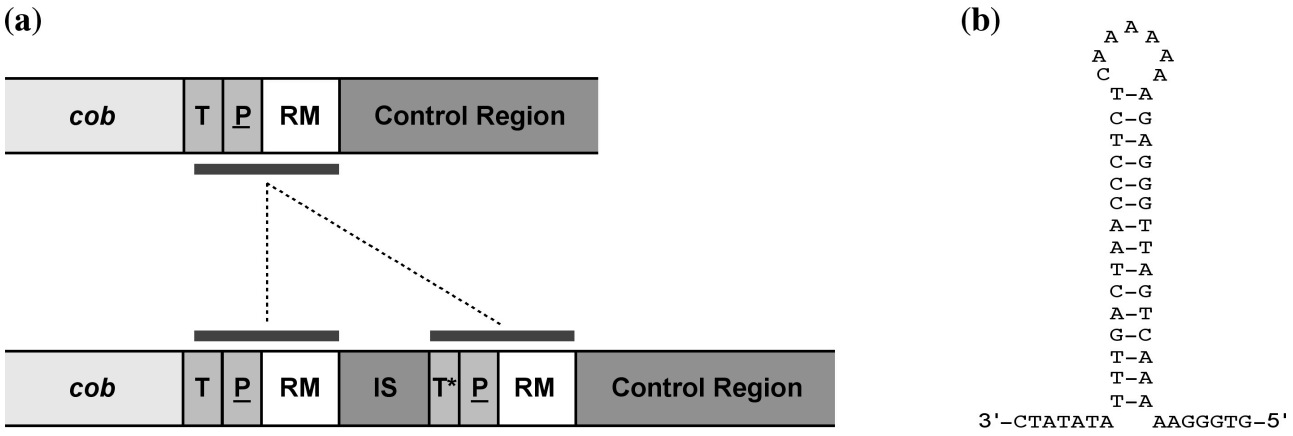
(Gymnophiona) based on complete mitochondrial genomes and nuclear RAG1. Mol. Phylogenet. Evol. 33:413-427.
San Mauro, D., D. J. Gower, R. Zardoya, and M. Wilkinson. 2006. A hotspot of gene order rearrangement by tandem duplication and random loss in the vertebrate mitochondrial genome. Mol. Biol. Evol. 23:227-234.
San Mauro, D., M. Vences, M. Alcobendas, R. Zardoya, and A. Meyer. 2005. Initial diversification of living amphibians predated the breakup of Pangaea. Am. Nat. 165:590-599.
Shimodaira, H. 2002. An approximately unbiased test of phylogenetic tree selection. Syst. Biol. 51:492-508.
Shimodaira, H., and M. Hasegawa. 2001. CONSEL: for assessing the confidence of phylogenetic tree selection. Bioinformatics 17:1246-1247.
Smith, A. G., D. G. Smith, and B. M. Funnell. 1994. Atlas of Mesozoic and Cenozoic coastlines. Cambridge University Press, Cambridge.
Springer, M. S., R. W. DeBry, C. J. Douady, H. M. Amrine, O. Madsen, W. W. deJong, and M. J. Stanhope. 2001. Mitochondrial versus nuclear gene sequences in deep-level mammalian phylogeny reconstruction. Mol. Biol. Evol. 18:132-143.
StatSoft Inc. 2001. STATISTICA (data analysis software system), version 6. http://www.statsoft.com.
Strimmer, K., and A. Rambaut. 2001. Inferring confidence sets of possible misspecified gene trees. Proc. R. Soc. London B 269:137-142.
Swofford, D. L. 1998. PAUP*: phylogenetic analysis using parsimony (*and other methods), version 4.0. Sinauer Associates, Inc., Sunderland, MA, USA.
Taylor, E. H. 1968. The Caecilians of the world: A taxonomic analysis. University of Kansas Press, Lawrence, KS.
Thompson, J. D., T. J. Gibson, F. Plewniak, J. Jeanmougin, and D. G. Higgins. 1997. The CLUSTAL X windows interface: Flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res. 25:4876-4882.
Thorne, J. L., and H. Kishino. 2002. Divergence time and evolutionary rate estimation with multilocus data. Syst. Biol. 51:689-702.
Thorne, J. L., H. Kishino, and I. S. Painter. 1998. Estimating the rate of evolution of the rate of molecular evolution. Mol. Biol. Evol. 15:1647-1657.
Wake, M. H. 1977. The reproductive biology of caecilians: an evolutionary perspective. Pages 73-101 *in* Reproductive biology of amphibians (E. H. Taylor, and S. I. Guttman, eds.). Plenum Press, New York.
Wilkinson, M. 1992. The phylogenetic position of the Rhinatrematidae (Amphibia: Gymnophiona): evidence from the larval lateral line system. Amphibia-Reptilia 13:74-79.
Wilkinson, M. 1996. The heart and aortic arches of rhinatrematid caecilians (Amphibia: Gymnophiona). Zoomorphology 105:277-295.
Wilkinson, M. 1997. Characters, congruence and quality: a study of neuroanatomical and traditional data in caecilian phylogeny. Biol. Rev. 72:423-470.
Wilkinson, M., S. P. Loader, D. J. Gower, J. A. Sheps, and B. L. Cohen. 2003. Phylogenetic relationships of African caecilians (Amphibia: Gymnophiona): insights from mitochondrial rRNA gene sequences. Afr. J. Herpetol. 52:83-92.
Wilkinson, M., and R. A. Nussbaum. 1996. On the phylogenetic position of the Uraeotyphlidae (Amphibia: Gymnophiona). Copeia 1996:550-562.
Wilkinson, M., and R. A. Nussbaum. 1997. Comparative morphology and evolution of the lungless caecilian *Atretochoana eiselti* (Taylor) (Amphibia: Gymnophiona: Typhlonectidae). Biol. J. Linn. Soc. 62:39-109.
Wilkinson, M., and R. A. Nussbaum. 1999. Evolutionary relationships of the lungless caecilian *Atretochoana eiselti* (Amphibia: Gymnophiona: Typhlonectidae). Zool. J. Linn. Soc. 126:191-223.
Wilkinson, M., and R. A. Nussbaum. 2006. Caecilian phylogeny and classification. Pages 39-78 *in* Reproductive biology and phylogeny of Gymnophiona (Caecilians) (J.-M. Exbrayat, ed.) Science Publishers, Enfield, NH.
Wilkinson, M., J. A. Sheps, O. V. Oommen, and B. L. Cohen. 2002. Phylogenetic relationships of Indian caecilians (Amphibia: Gymnophiona) inferred from mitochondrial rRNA gene sequences. Mol. Phylogenet. Evol. 23:401-407.
Wolstenholme, D. R. 1992. Animal mitochondrial DNA: structure and evolution. Int. Rev. Cytol. 141:173-216.
Wong, T. W., and D. A. Clayton. 1985. In vitro replication of human mitochondrial DNA: accurate initiation at the origin of light-strand synthesis. Cell 42: 951-958.
Yang, Z. 1994. Maximum likelihood phylogenetic estimation from DNA sequences with variable rates over sites: approximate methods. J. Mol. Evol. 39:306-314.
Yang, Z. 1997. PAML: a program package for phylogenetic analysis by maximum likelihood. Comput. Appl. Biosci. 13:555-556.
Yang, Z. 1998. On the best evolutionary rate for phylogenetic analysis. Syst. Biol. 47:125-133.
Zardoya, R., E. Malaga-Trillo, M. Veith, and A. Meyer. 2003. Complete nucleotide sequence of the mitochondrial genome of a salamander, *Mertensiella luschani*. Gene 317:17-27.
Zardoya, R., and A. Meyer. 1996. Phylogenetic performance of mitochondrial protein-coding genes in resolving relationships among vertebrates. Mol. Biol. Evol. 13:933-942.
Zardoya, R., and A. Meyer. 2000. Mitochondrial evidence on the phylogenetic position of Caecilians (Amphibia: Gymnophiona). Genetics 155:765-775.
Zwickl, D. J., and D. M. Hillis. 2002. Increased Taxon Sampling Greatly Reduces Phylogenetic Error. Syst. Biol. 51:588-598.

APPENDIX

DISTINCT STRUCTURAL FEATURES OF THE MT GENOMES OF *BOULENGERULA TAITANUS* AND *GEOTRYPETES SERAPHINI*

The mt genome of *B. taitanus* conforms to the vertebrate consensus mt gene arrangement (Jameson et al., 2003), but possesses a duplicated region comprising the 3′-half portion of *trnT*, *trnP*, and a 50-bp-long motif of the 5′-part of the control region (Appendix Fig. 1a). The duplicated portions are separated by a 61-bp-long non-coding spacer, and are virtually identical in nucleotide sequence. Since repetitive sequences are thought to be lost rapidly from the mtDNA under strong selective pressure to constrain its size and gene number (Wolstenholme, 1992), there can be two alternative explanations for the high sequence conservation in the two duplicated portions: (1) the duplication took place not long ago (in evolutionary time); or (2) the duplication occurred long ago, but both portions followed concerted evolution after the duplication event (Kumazawa et al., 1996). According to the model of mt gene rearrangement by tandem duplication followed by random loss of redundant genes (e.g., Moritz and Brown, 1987; Moritz et al., 1987; San Mauro et al., 2006), this duplication may constitute an intermediate step of the rearrangement process (Macey et al., 1997). In such case, the 61-bp-long non-conserved spacer separating both duplicated portions should be part of the control region (following downstream the 50-bp-long motif) that either has evolved faster than the duplicated portion or has not followed concerted evolution. In contrast, if this spacer is actually not part of the control region, alternative explanations for this duplication event, such as transposition (Macey et al., 1997) or intramolecular recombination (Lunt and Hyman, 1997), cannot be completely ruled out. Duplications involving the same mt region have been reported for other vertebrates, such as the amphisbaenian *Bipes biporus* (Macey et al., 1998) and the scincomorph lizard *Cordylus warreni* (Kumazawa, 2004), providing further evidence that duplications more likely occur in close proximity to (or involving) replication origins (e.g., Kumazawa et al., 1998; Mindell et al., 1998; Moritz and Brown, 1987).

The mt genome organization of *G. seraphini* also conforms to the vertebrate consensus mt gene arrangement (Jameson et al., 2003), but lacks the origin of light-strand replication (O_L) on its typical position. In the mtDNA of this caecilian, only five nucleotides remain between *trnN* and *trnC*, and no stem-loop structure can be identified. However, the mt genome of this caecilian possesses a long non-coding intergenic spacer (301 bp) between *trnI* and *trnQ* where there is a 37-bp-long region that can be folded into a stem-loop structure (Appendix Fig. 1b). The nucleotide sequence of this stem-loop structure is fairly dissimilar to those reported for other caecilian O_L's (San Mauro et al., 2004b; San Mauro et al., 2006; Zardoya and Meyer, 2000), and lacks some functional motifs identified in human and mouse O_L's as necessary for light strand replication (Brennicke and Clayton, 1981; Hixson et al., 1986; Wong and Clayton, 1985). However, in absence of a functional O_L on its typical position, this stem-loop structure may have been co-opted for the function of light strand replication, as it has been reported for some tRNA genes in other animal species (Clary and Wolstenholme, 1985; Clayton, 1982). A long intergenic spacer between *trnI* and *trnQ* has been also reported in the mt genome of the snake *Dinodon semicarinatus* (Kumazawa et al., 1998), which corresponds to a duplicated control region, □*trnP*, and *trnL1*. However, unlike *G. seraphini*, the mt genome of that snake does possess a functional O_L. Furthermore, sequence similarity between the spacer found in *G. seraphini* and any part of its mtDNA control region is low, and BLAST searches (Altschul et al., 1990) produced no close matches.



APPENDIX FIGURE 1. (a) Genomic duplication found in *Boulengerula taitanus*. tRNA genes are abbreviated by the corresponding one-letter amino acid code, and genes encoded by the light-strand are underlined. IS, intergenic spacer; RM, repeated motif; T*, truncated *trnT* (27 bp shorter). (b) Proposed secondary structure of the stem-loop region found in *Geotrypetes seraphini* between *trnI* and *trnQ*.

7. DISCUSSION

7.1. *Origin and Diversification of Living Caecilians*

The origin and divergence of the three orders of living amphibians has long remained a hotly debated topic in vertebrate evolution. According to the results presented in Publication I of this thesis, the separation of the three lissamphibian orders occurred during the Late Paleozoic (between 417 and 317 million years ago [mya]), when the ancestral lineage of caecilians separated from the common ancestor of Batrachia, and, shortly thereafter, frogs diverged from salamanders. Even analyses with alternative and single calibrations yielded a Paleozoic age of separation among the three orders of living amphibians, although estimated time frames were younger in some cases (between 392 and 305 mya). This is in agreement with two other recent studies based on mitogenomic (Zhang et al., 2005) and nuclear-mitochondrial combined (Roelants et al., in ms.) evidence that also situate the separation of the three orders of living amphibians in the Late Paleozoic (between 353 and 312 mya).

All these evidences may indicate that the separation of the three orders of modern amphibians in the Paleozoic (most likely during the Carboniferous) occurred almost immediately (in evolutionary time) after the "jump to land" of sarcopterygian fishes (over 360 mya; Benton, 2005; Carroll, 1988), as postulated by Benton (1990), Milner (1993) and Carroll et al. (2004), and in parallel to the diversification of extinct lineages of amphibians (such as *Acanthostega* or *Ichthyostega*). Such a rapid radiation event may be the cause for the lack of fossils that represent plausible ancestors or morphological immediate sister taxa of all three lissamphibian orders, and the particularly short branch lengths connecting the nodes among them, thereby rendering phylogenetic inference more difficult. Furthermore, the estimated Paleozoic separation of the three orders of living amphibians predating Pangaea fragmentation (over 180 mya; Gurnis, 1988; Smith et al., 1994) invalidate the earlier hypothesis that salamanders and caecilians arose in the Mesozoic from a common ancestor by vicariance directly linked to the breakup of Pangaea (Feller and Hedges, 1998). Unfortunately, the relatively large confidence intervals of the age estimates do not shed any new light on the problematic ancestry of modern amphibians. Zhang et al. (2005) indicated that their divergence time estimates provided support for a temnospondyl ancestry of lissamphibians, but this was recently criticized by Lee and Anderson (2006) who thoroughly

argued that Zhang et al.'s (2005) conclusions were poorly supported. In general, and given the virtually overlapping stratigraphic ranges of the Temnospondyli and the Lepospondyli (Heatwole and Carroll, 2000; Milner, 1993), determining which of these two extinct groups (or both) are the ancestors of the living amphibians remains essentially a paleontological enterprise.

The results presented in Publication I, based exclusively on *rag1*, indicate that the time of splitting of the most ancient clades of modern caecilians (Rhinatrematidae, and the Ichthyophiidae + Uraeotyphlidae clade) occurred during the Mesozoic, between 256 and 160 mya, before the fragmentation of Gondwana (130 to 86 mya; Pitman III et al., 1993; Rabinowitz et al., 1983). However, basal caecilian divergences were estimated to be younger when using alternative single calibrations (means of the Rhinatrematidae split ranging from 177 to 150 mya). Analyses of the more comprehensive dataset of Publication IV (mitogenomic + *rag1*) also yielded younger age estimates, and with smaller confidence intervals, that indicate that the caecilian crown-group originated in the Middle-Late Jurassic (split of Rhinatrematidae 166 to 148 mya), and that all subsequent divergences of caecilian lineages (of the ones used in the study) occurred during the Cretaceous (between 138 and 66 mya). These ages imply that the divergence of higher caecilian lineages was concomitant with the fragmentation of Gondwana.

These results are congruent with those of the recent study by Roelants et al. (in ms.), although their initial splits for crown-group caecilians (those of Rhinatrematidae, and the Ichthyophiidae + Uraeotyphlidae clade) are slightly older than the results presented in Publication IV (195-132 versus 166-132 mya). Another previous study (Wilkinson et al., 2002) recovered age estimates for caecilian divergences more similar to those oldest presented in Publication I, suggesting that the initial crown-group split predated Pangaea fragmentation, and that even major higher caecilian splits predated Gondwana fragmentation. This latter scenario allows a straightforward explanation of the current distribution of higher caecilians in South America, Africa, India, and the Seychelles, as well as of the putative African affinities of a Paleocene caeciliid-like fossil (*Apodops pricei*) found in South America (Estes and Wake, 1972): the ancestors of major higher caecilian lineages were already distributed throughout Gondwana before its initial fragmentation (Duellman and Trueb, 1994). In contrast, the younger age estimates recovered in Publication I, those of Publication IV, and those by Roelants et al. (in ms.) imply some sort of dispersion (most likely through land bridges connecting the western part of Gondwana [Africa and South America] with landmasses of the disintegrating eastern Gondwana [India, Madagascar, and the Seychelles])

to account for the current distribution of some higher caecilians, particularly Indian and Seychellean caeciliids. Some palaeogeological models incorporating land bridges between western and eastern Gondwanan landmasses during the late Mesozoic have been recently proposed, and have helped explaining the puzzling affinities and distribution of several vertebrate groups including dinosaurs, frogs, lizards, and mammals (Briggs, 2003; Chatterjee and Scotese, 1999).

7.2. *Caecilian Phylogeny*

All relevant analyses in Publication II provide strong support for the Batrachia hypothesis (frogs as sister group of salamanders, to the exclusion of caecilians; Milner, 1988) in agreement with most morphological (Duellman and Trueb, 1994; McGowan and Evans, 1995; Milner, 1988; Rage and Janvier, 1982; Trueb and Cloutier, 1991), and the most recent molecular (Frost et al., 2006; Roelants et al., in ms.; Zardoya and Meyer, 2001; Zhang et al., 2005) studies. Also, the Batrachia hypothesis is the best explanation in Publication I (albeit not strongly supported). All this evidence contradicts early molecular studies based on ribosomal genes (Feller and Hedges, 1998; Hay et al., 1995; Hedges and Maxson, 1993; Hedges et al., 1990; Larson and Wilson, 1989) that supported a sister group relationship of caecilians and salamanders. In any case, and despite increasing evidence supporting the Batrachia hypothesis, this particular node of the amphibian phylogenetic tree appears to be remarkably challenging, both for molecular and morphological data (see Publication I, and Schoch and Milner, 2004).

Within Gymnophiona, results from Publication II strongly support the conventional view based on morphology (Nussbaum, 1977; Nussbaum, 1979; Wilkinson, 1992; Wilkinson, 1996; Wilkinson, 1997) and molecules (Frost et al., 2006; Hedges et al., 1993; Roelants et al., in ms.) that the Rhinatrematidae is the sister group of all other extant caecilians. The sister group relationship of Ichthyophiidae and Uraeotyphlidae, and the monophyly of higher caecilians are also strongly supported by the results in Publications II and IV (both of them based on mitogenomic + *rag1* data). In general, all inferred phylogenetic relationships outside the higher caecilians are strongly supported (see Fig. 7 in Publication II, and Fig. 1 in Publication IV), and in full agreement with the most recent molecular (Frost et al., 2006; Roelants et al., in ms.; Wilkinson et al., 2003; Wilkinson et al., 2002) and morphological (Wilkinson, 1997; Wilkinson and Nussbaum, 1996; Wilkinson and Nussbaum, 1999) studies,

and the current caecilian classification (Wilkinson and Nussbaum, 2006). Also, the analyses of Publication I recover fully congruent phylogenetic relationships. It has been suggested (Carroll, 2000a; Duellman and Trueb, 1994; Nussbaum and Wilkinson, 1989) that these phylogenetic pathways may reflect the successive complex specializations that caecilians evolved in morphology (such as skull reinforcement and tail loss) and life history (such as direct development or viviparity) for their terrestrial burrowing habits (not including typhlonectids that are secondarily adapted to aquatic or semi-aquatic habitats; Taylor, 1968; Wilkinson and Nussbaum, 1999).

Unlike the general agreement outside the higher caecilians, there is a lot more uncertainty on the inter- and intrafamilial phylogenetic relationships within the higher caecilians (outlined in Publications II and IV, and also in Wilkinson, 1997; Wilkinson et al., 2003; Wilkinson and Nussbaum, 2006). Indeed, the most comprehensive morphological study to date (Wilkinson, 1997) was unable to conclusively resolve relationships among these higher caecilians, and only two recent molecular studies (Frost et al., 2006; Roelants et al., in ms.) have provided a supported (but conflicting with each other) picture of the phylogenetic relationships within this clade. The phylogenetic results in Publication IV based on mitogenomic + *rag1* data are in full agreement with those by Roelants et al. (in ms.) that reflect a perhaps more traditional assemblage, with Scolecomorphidae recovered as the sister group of all other higher caecilians (Duellman and Trueb, 1994; Nussbaum and Wilkinson, 1989), and with Caeciliidae recovered as paraphyletic with respect to Typhlonectidae (Hedges et al., 1993; Nussbaum, 1979; Wilkinson, 1997; Wilkinson et al., 2003) (see Fig. 1 in Publication IV).

In contrast to this scenario, Frost et al.'s (2006) study indicated that Caeciliidae is paraphyletic with respect to Typhlonectidae as well as with respect to Scolecomorphidae. This hypothesis had already been suggested (albeit only tentatively) by a previous molecular study (Wilkinson et al., 2003), and involves an alternative branching scenario at the base of the higher caecilian tree (with a *Boulengerula* + *Herpele* clade recovered as the sister group of all other higher caecilians). Interestingly, the only node that is not confidently supported by the two methods of phylogenetic inference used in Publication IV is the one of the basal split within the higher caecilians (node D in Fig. 1 of Publication IV), involving the phylogenetic position of *Scolecomorphus* and *Boulengerula*. Furthermore, phylogenetic reconstructions using the most informative subset of the full mitogenomic + *rag1* dataset (see Publication IV) recover *Boulengerula* (and not *Scolecomorphus*) at the base of the higher caecilian tree. Indeed, this latter branching arrangement cannot be rejected by the approximately unbiased

test, although it is strongly rejected by parametric bootstrapping analyses (see Table 3 in Publication IV). Hence, the caecilian tree shown in Fig. 1 of Publication IV (with *Scolecomorphus* as the sister taxon of all other higher caecilians) appears to be the most robust reconstruction of the caecilian phylogeny given the available mitogenomic + *rag1* data, but some uncertainty regarding the position of *Scolecomorphus* and *Boulengerula* cannot be completely ruled out. According to the information analyses performed in Publication IV, stabilization of this node could potentially be achieved by targeting additional taxa at the sister positions of *Scolecomorphus*, *Boulengerula*, and *Rhinatrema* (most likely *Crotaphatrema*, *Herpele*, and *Epicrionops*, respectively; Frost et al., 2006; Roelants et al., in ms.; Wilkinson and Nussbaum, 2006) in future studies on caecilian phylogenetics.

7.3. *The Caecilian Mitochondrial Genome*

The mitochondrial DNA of all nine caecilian species studied (namely, *Rhinatrema bivittatum*, *Ichthyophis glutinosus*, *Uraeotyphlus* cf. *oxyurus*, *Scolecomorphus vittatus*, *Boulengerula taitanus*, *Typhlonectes natans*, *Gegeneophis ramaswamii*, *Geotrypetes seraphini*, and *Siphonops annulatus*; see Publications II, III, and IV, and Zardoya and Meyer [2000]) is a circular molecule averaging 16000 bp that, in general, conforms to the vertebrate consensus mitogenomic organization (Fig. 8) (but see exceptions below). As in most other amphibians (e.g., Roe et al., 1985; San Mauro et al., 2004; Zardoya et al., 2003; Zhang et al., 2003a), all caecilian mitochondrial genomes encode for two ribosomal RNAs, 22 tRNAs (that can be typically folded into canonical cloverleaf secondary structures), and 13 protein-coding genes, with the exception of *Gegeneophis ramaswamii*'s lack of *trnF*, and *Boulengerula taitanus*' duplication of *trnP* (see below). Overall base compositions are biased against guanine in all caecilian mitochondrial genomes. This is a typical feature of the vertebrate mitochondrial DNA, and is mainly due to a strong selection against the use of guanine at the third codon positions of the protein-coding genes (Zardoya and Meyer, 2000). As described for other animal mitochondrial genomes (Zardoya and Meyer, 2001), the highest sequence variability of caecilian mitochondrial DNA is mainly detected in non-coding regions, in the dihydrouridine and T□C loops of tRNA genes, in the 5' and 3' ends of protein-coding genes, and in several highly variable regions of ribosomal genes, which suggests more relaxed selection on these particular mitogenomic regions.

In most caecilians, the mitochondrial control region possess three conserved blocks (CSB-1, CSB-2, and CSB-3; Walberg and Clayton, 1981) at the 3'-end, and two polypyrimidine tracts, PP-1 and PP-2, upstream from the CSB-2 and CSB-3 motifs (see Fig. 2 in Publication II) that are likely involved in regulatory aspects of the origin of the heavy-strand replication. Termination-associated sequences (TAS, which have a putative role in arresting replication; Doda et al., 1981) and tandem repeats (only reported for *Typhlonectes natans*; Zardoya and Meyer, 2000) are typically absent from the caecilian mitochondrial control region. As in most vertebrates, the putative origin of light-strand replication (O_L) of the caecilian mitochondrial genome is located within the WANCY tRNA cluster (Boore, 1999; Jameson et al., 2003; Seutin et al., 1994) (see Fig. 1 in Publication II), between *trnN* and *trnC* (but see exceptions below), and has the potential to fold into a stem-loop secondary structure, sharing some nucleotides with the flaking tRNAs (see Fig. 3 in Publication II). As described for human mitochondrial DNA (Wong and Clayton, 1985), light-strand synthesis is probably initiated in a stretch of thymines in the O_L loop (see Fig. 3 in Publication II, and Fig. 4 in Publication III). The 5'-GCCGG-3' motif that in human mitochondrial DNA is involved in the transition from RNA synthesis to DNA synthesis (Hixson et al., 1986) is generally not conserved in caecilians (see Fig. 3 in Publication II).

Distinct structural features are found in the mitochondrial genomes of *Gegeneophis ramaswamii*, *Siphonops annulatus*, *Boulengerula taitanus*, and *Geotrypetes seraphini*. Moreover, the mitochondrial genomes of *Rhinatrema bivittatum* and *Uraeotyphlus* cf. *oxyurus* have unusually long intergenic spacers between *trnT* and *trnP* (see Publication II). Other cases of long intervening non-coding sequences have been reported in some salamanders (Zardoya et al., 2003; Zhang et al., 2003b), but in all cases sequence similarities are low, suggesting that they have independent origins (likely related to duplication or transposition events; Macey et al., 1997; Moritz and Brown, 1986; Moritz and Brown, 1987).

The mitochondrial genome of *Gegeneophis ramaswamii* lacks the *trnF* gene (see Publication II). In this species, the 3' end of *rrnS* follows the 5' end of the control region, and *trnF* is not found elsewhere in the genome. This presumably derived absence is unique among known vertebrate mitochondrial genomes. Absence of other tRNA genes has been previously reported in marsupials (Janke et al., 2002; Janke et al., 1997), and the tuatara (Rest et al., 2003). In marsupials, it has been shown that an alternative tRNA (for lysine) of nuclear origin is imported into mitochondria to participate in the translation process (Dorner et al., 2001). Given that the usage of phenylalanine in the mitochondrial proteins of *Gegeneophis*

ramaswamii is comparable to that in the other caecilians, an analogous importation may be implicated.

In *Siphonops annulatus* mitochondrial DNA, the WANCY genomic region is rearranged departing from the consensus gene organization found in vertebrates (Boore, 1999; Jameson et al., 2003; Seutin et al., 1994), and all other studied caecilians. Two other species of *Siphonops* (*Siphonops paulensis* and *Siphonops hardyi*) also possess the same WANCY region gene orders, and the closely related (Taylor, 1968; Wilkinson and Nussbaum, 1992) monotypic *Luetkenotyphlus brasiliensis* possesses another WANCY region gene arrangement, different from both the vertebrate consensus and that of *Siphonops* species (see Figs. 1 and 2 in Publication III). Given that duplications of genes appear to be infrequent among mitochondrial genomes (Boore, 2000), independent duplications of the WANCY region in *Luetkenotyphlus* and in *Siphonops* provide a less plausible explanation of the derived gene orders of these closely related caecilians than their resulting from a single ancestral tandem duplication of the entire WANCY region followed by almost instant loss of two redundant gene duplicates (*trnW*, *trnC*), and independent, random loss of three (*trnA*, *trnN*, *trnY*) redundant gene duplicates in *Siphonops* and *Luetkenotyphlus* (see Fig. 1 in Publication III). An alternative reconstruction in which all redundant duplicates are independently lost after the first speciation event (the split between *Siphonops* and *Luetkenotyphlus*) seems equally plausible. In all *Siphonops* and *Luetkenotyphlus*, there are intergenic spacers (4-13 bp long) at positions expected of pseudogenes under the tandem duplication – random loss (TDRL) model (Boore, 2000; Moritz and Brown, 1987; Moritz et al., 1987) (see Fig. 2 in Publication III). A more substantial intergenic spacer between the *trnA* gene and the O_L is similar to the known, functional *trnN* genes of caecilians, but with substantial length and substitution mutations (see Fig. 3 in Publication III), and can be more confidently identified as the *trnN* pseudogene (Ψ*trnN*) predicted by the TDRL model. Although their anticodon sequences are conserved, the *Siphonops* and *Luetkenotyphlus* Ψ*trnN* have all lost the potential to fold into stable cloverleaf structures (indicating loss of primary function), and have evolved more than twice as fast as their functional paralogs (see Publication III), suggesting that, following duplication, the redundant *trnN* paralogs have experienced more relaxed selective constraints (Moritz and Brown, 1987). Interestingly, the derived WANCY region of *Siphonops* is exceptionally similar to that of marsupials (Pääbo et al., 1991), and the gene arrangement of functional tRNA genes is identical, providing an example of convergent derived gene order. As discussed in Publication III, this convergence, together with comparative data for other available vertebrate complete mitochondrial genomes, provide further evidence that the

WANCY genomic region is a hotspot for gene order change (where convergent gene rearrangements are not unlikely), and support the view that TDRL is the dominant mechanism of gene order rearrangement in vertebrate mitochondrial genomes.

The mitochondrial genome of *Boulengerula taitanus* possesses a duplicated region comprising the 3' half portion of *trnT*, *trnP*, and a 50 bp long motif of the 5' part of the control region (see Appendix Figure 1 in Publication IV). The duplicated portions are separated by a 61 bp long non-coding spacer, and are virtually identical in nucleotide sequence. Since repetitive sequences are thought to be lost rapidly from the mitochondrial DNA under strong selective pressure to constrain its size and gene number (Wolstenholme, 1992), there can be two alternative explanations for the high sequence conservation in the two duplicated portions: that the duplication took place not long ago (in evolutionary time); or that the duplication occurred long ago, but both portions followed concerted evolution after the duplication event (Kumazawa et al., 1996). According to the TDRL model of mitochondrial gene rearrangement (Boore, 2000; Moritz and Brown, 1987; Moritz et al., 1987), this duplication may constitute an intermediate step of the rearrangement process (Macey et al., 1997). In such case, the 61 bp long non-conserved spacer separating both duplicated portions should be part of the control region (following downstream the 50 bp long motif) that either has evolved faster than the duplicated portion or has not followed concerted evolution. In contrast, if this spacer is actually not part of the control region, alternative explanations for this duplication event, such as transposition (Macey et al., 1997) or intramolecular recombination (Lunt and Hyman, 1997), cannot be completely ruled out.

Finally, the mitochondrial DNA of *Geotrypetes seraphini* lacks the O_L on its typical position. In the mitochondrial genome of this caecilian, only five nucleotides remain between *trnN* and *trnC*, and no stem-loop structure can be identified. However, the mitochondrial genome of this caecilian possesses a long non-coding intergenic spacer (301 bp) between *trnI* and *trnQ* where there is a 37 bp long region that can be folded into a stem-loop structure (see Appendix Figure 1 in Publication IV). The nucleotide sequence of this stem-loop structure is fairly dissimilar to those reported for other caecilian O_L (see Publication II), and lacks some functional motifs identified in human and mouse O_L as necessary for light strand replication (Brennicke and Clayton, 1981; Hixson et al., 1986; Wong and Clayton, 1985). However, in absence of a functional O_L on its typical position, this stem-loop structure may have been co-opted for the function of light strand replication, as it has been reported for some tRNA genes in other animal species (Clary and Wolstenholme, 1985; Clayton, 1982).

7.4. *Phylogenetic Utility of Mitochondrial and rag1 Data*

Substitution rate of *rag1* at both nucleotide and amino acid levels is relatively slower than that of almost all mitochondrial genes (see Fig. 4 in Publication II, and Fig. 2 in Publication IV), being similar to those of the most conservative mitochondrial protein-coding genes (cytochrome c oxidase subunits; Lopez et al., 1997; Russo et al., 1996; Zardoya and Meyer, 1996b), and in agreement with previous studies that have indicated the relatively slow evolution of nuclear *rag1* (Groth and Barrowclough, 1999). Most other mitochondrial genes examined (ribosomal, tRNA, and other protein-coding at the amino acid level or without the third position) show a moderately faster rate of evolution. In contrast, the rate of evolution of third codon positions of mitochondrial protein-coding genes is extremely fast (over 100-fold faster) compared to those of all other datasets analysed (see Fig. 2 in Publication IV), which agrees with previous studies that reported the faster evolutionary rates of third codon positions with respect to first and second positions (Irwin et al., 1991; Johnson and Sorenson, 1998; Li and Graur, 1991). This extremely fast substitution rate makes phylogenetic signal to be eroded due to accumulation of superimposed, homoplasic nucleotide substitutions, and it is the main reason why most studies (including the ones presented in this thesis) on deep vertebrate divergences use mitochondrial protein-coding genes either at the amino acid level or after exclusion of third codon positions (even complex models of sequence evolution can hardly account for the extreme variation at these positions).

The assessment of phylogenetic information using Goldman's (1998b) method (see Publication IV) reveals that the most informative genes are those encoding for tRNAs (all concatenated) and ribosomal subunits, some genes encoding subunits of the NADH dehydrogenase complex (*nad2*, *nad5*, and *nad6*; all without the third position), and the nuclear slower-evolving *rag1*, this latter one particularly in the more internal branches of the caecilian tree. The phylogenetic performance of most of these genes is well known, and they have long been indicated as good or adequate molecular markers for inferring ancient vertebrate phylogenies (Cummings and Meyer, 2005; Cummings et al., 1995; Groth and Barrowclough, 1999; Kumazawa and Nishida, 1993; Martin, 1999; Mindell and Honeycutt, 1990; Miya and Nishida, 2000; Mueller, 2006; Murphy et al., 2001; Russo et al., 1996; Zardoya and Meyer, 1996b). Overall mitogenomic phylogenetic information (calculated as the combination of all mitochondrial genes) is higher than that of any single mitochondrial gene (although the tRNA genes account for a large part of the mitogenomic information; see Publication IV) and that of nuclear *rag1*. This is in agreement with many studies that have

highlighted the phylogenetic utility of complete mitochondrial genomes in studying deep vertebrate divergences (e.g., Arnason et al., 2002; Mindell et al., 1999; Miya et al., 2003; Zardoya and Meyer, 1996a; Zardoya and Meyer, 1998; Zardoya and Meyer, 2001). Moreover, whole mitogenomic data makes available additional information (other than just that of sequence characters), such as gene order, that in some cases (when rearranged) may provide evidence of phylogenetic relationships, and may help to better understand the mechanisms driving the evolution of the mitochondrial genome (see Publication III).

Given the above mentioned evolutionary rates and phylogenetic information, both mitochondrial genomes (with protein-coding genes at the amino acid level, or without the third position) and nuclear *rag1* are potentially useful molecular markers for the study of deep caecilian divergences. In general, both types of data showed high phylogenetic performance and provided well-supported resolution of caecilian phylogenetic relationships in the studies presented in this thesis.

8. CONCLUSIONS

From the studies presented in this Ph.D. thesis aiming to study the phylogeny and molecular evolution of caecilian amphibians, the following conclusions can be drawn:

1. The ancestral lineage of caecilians separated from the common ancestor of Batrachia (frogs and salamanders) during the Late Paleozoic (Carboniferous), before the fragmentation of Pangaea.

2. The caecilian crown-group originated in the Middle Mesozoic (Jurassic), before the breakup of Gondwana, and subsequent divergences of major caecilian lineages occurred during the Late Mesozoic (Cretaceous).

3. Rhinatrematidae is the sister group of all other extant caecilians, followed by an Ichthyophiidae + Uraeotyphlidae clade. The remaining three families (higher caecilians) constitute a monophyletic group, with Scolecomorphidae as the sister group of a paraphyletic Caeciliidae with respect to Typhlonectidae. The paraphyly of Caeciliidae with respect to Scolecomorphidae cannot be completely ruled out.

4. The caecilian mitochondrial genome conforms in general to the vertebrate consensus mitogenomic organization. However, distinct structural features are found in the mitochondrial genomes of *Gegeneophis ramaswamii* (lack of the *trnF* gene), *Siphonops annulatus* (gene order rearrangement of the WANCY region), *Boulengerula taitanus* (genomic duplication adjacent to the control region), and *Geotrypetes seraphini* (lack of the O_L on its typical position, but perhaps displaced between *trnI* and *trnQ*).

5. The derived gene order of the genomic WANCY region of *Siphonops* and *Luetkenotyphlus*, together with comparative data for other vertebrate complete mitochondrial genomes, support the view that tandem duplication followed by random loss of redundant genes is the dominant mechanism of gene order rearrangement in

vertebrate mitochondrial genomes, and provide evidence that the WANCY region is a hotspot for gene order change.

6. Both complete mitochondrial genomes and the nuclear *rag1* gene show a high phylogenetic performance and are potentially useful molecular markers for the study of deep caecilian divergences.

7. Future phylogenetic studies tackling the controversial paraphyly of Caeciliidae with respect to Scolecomorphidae could greatly benefit from the combined use of highly informative genes (such as those for mitochondrial tRNAs, ribosomal subunits, and some subunits of the NADH dehydrogenase complex; and nuclear *rag1*), as well as the addition of taxa at the sister positions of *Scolecomorphus*, *Boulengerula*, and *Rhinatrema*.

9. CONCLUSIONES (Spanish translation of Conclusions)

De los estudios presentados en esta Tesis Doctoral sobre filogenia y evolución molecular en cecilias, pueden extraerse las siguientes conclusiones:

1. El linaje ancestral de las cecilias se separó del ancestro común de Batrachia (ranas y salamandras) durante el Paleozoico superior (Carbonífero), antes de la fragmentación de Pangea.

2. Las cecilias modernas se originaron a mediados del Mesozoico (Jurásico), antes de la rotura de Gondwana, y las divergencias subsecuentes de los principales linajes de cecilias ocurrieron durante el Mesozoico superior (Cretácico).

3. Rhinatrematidae es el grupo hermano de todas las demás cecilias modernas, seguido del clado Ichthyophiidae + Uraeotyphlidae. Las tres familias restantes (cecilias superiores) constituyen un grupo monofilético, con Scolecomorphidae como grupo hermano de Caeciliidae, que aparece parafilética con respecto a Typhlonectidae. La parafilia de Caeciliidae con respecto a Scolecomorphidae no puede ser completamente descartada.

4. El genoma mitocondrial de las cecilias tiene en general la organización mitogenómica consenso de vertebrados. Sin embargo, se encuentran características estructurales distintivas en los genomas mitocondriales de *Gegeneophis ramaswamii* (ausencia del gen *trnF*), *Siphonops annulatus* (reordenación de genes de la región WANCY), *Boulengerula taitanus* (duplicación genómica junto a la región control) y *Geotrypetes seraphini* (ausencia del O_L en su posición típica, con posible desplazamiento entre los genes *trnI* y *trnQ*).

5. El orden de genes derivado en la región genómica WANCY de *Siphonops* y *Luetkenotyphlus*, junto con datos comparativos para otros genomas mitocondriales completos de vertebrados, apoyan la idea de que la duplicación en tándem seguida de pérdida aleatoria de genes es el mecanismo dominante de reordenación de genes en el

genoma mitocondrial de vertebrados, y proporcionan evidencia de que la región genómica WANCY es un punto con alta tasa de reordenación de genes.

6. Tanto los genomas mitocondriales completos como el gen nuclear *rag1* muestran un alto rendimiento filogenético, y son marcadores moleculares potencialmente útiles para el estudio de divergencias antiguas de cecilias.

7. Estudios futuros sobre la controvertida parafilia de Caeciliidae con respecto a Scolecomorphidae podrían beneficiarse enormemente del uso combinado de genes altamente informativos (tales como los mitocondriales para los tRNAs, las subunidades ribosomales, y algunas subunidades del complejo NADH deshidrogenasa; y el nuclear *rag1*), así como la adición de taxones en la posición hermana de *Scolecomorphus*, *Boulengerula*, y *Rhinatrema*.

10. LITERATURE CITED

Adachi, J., and M. Hasegawa. 1996. Model of amino acid substitution in proteins encoded by mitochondrial DNA. J. Mol. Evol. 42:459-468.

Agrawal, A., Q. M. Eastman, and D. G. Schatz. 1998. Implications of transposition mediated by V(D) J-recombination proteins RAG1 and RAG2 for origins of antigen-specific immunity. Nature 394:744-751.

Akaike, H. 1973. Information theory as an extension of the maximum likelihood principle*in* Second international symposium of information theory (B. N. Petrov, and F. Csaki, eds.). Akademiai Kiado, Budapest, Hungary.

Alfaro, M. E., S. Zoller, and F. Lutzoni. 2003. Bayes or bootstrap? A simulation study comparing the performance of Bayesian Markov chain Monte Carlo sampling and bootstrapping in assessing phylogenetic confidence. Mol. Biol. Evol. 20:255-256.

Arnason, U., J. A. Adegoke, K. Bodin, E. W. Born, Y. B. Esa, A. Gullberg, M. Nilsson, R. V. Short, X. Xu, and A. Janke. 2002. Mammalian mitogenomic relationships and the root of the eutherian tree. Proc. Natl. Acad. Sci. USA 99:8151-8156.

Avise, J. C. 1994. Molecular Markers, Natural History, and Evolution. Chapman & Hall, New York.

Benton, M. J. 1990. Phylogeny of the major tetrapod groups: Morphological data and divergence dates. J. Mol. Evol. 30:409-424.

Benton, M. J. 2005. Vertebrate palaeontology, Third edition. Blackwell Publishing Ltd., Malden, MA.

Biju, S. D., and F. Bossuyt. 2003. New frog family from India reveals an ancient biogeographical link with the Seychelles. Nature 425:711-714.

Bolt, J. R. 1991. Lissamphibian origins. Pages 194-222 *in* Origins of the major groups of tetrapods: Controversies and consensus (H. P. Schultze, and L. Trueb, eds.). Cornell University Press, Ithaca.

Boore, J. L. 1999. Animal mitochondrial genomes. Nucleic Acids Res. 27:1767-1780.

Boore, J. L. 2000. The duplication/random loss model for gene rearrangement exemplified by mitochondrial genomes of deuterostome animals. Pages 133–147 *in* Comparative Genomics, Computational Biology Series, vol. 1 (D. Sankoff, and J. Nadeau, eds.). Kluwer Academic Publishers, Dordrecht, Netherlands.

Brennicke, A., and D. A. Clayton. 1981. Nucleotide assignment of alkalisensitive sites in mouse mitochondrial DNA. J. Biol. Chem. 256:10613–10617.

Bridge, D., C. W. Cunningham, B. Schierwater, R. DeSalle, and L. W. Buss. 1992. Classlevel relationships in the phylum Cnidaria: evidence from mitochondrial genome structure. Proc. Natl. Acad. Sci. USA 89:8750–8753.

Briggs, J. C. 2003. The biogeographic and tectonic history of India. J. Biogeogr. 30:381-388.

Bromham, L., and D. Penny. 2003. The modern molecular clock. Nat. Rev. Genet. 4:216-224.

Bruno, W. J., and A. L. Halpern. 1999. Topological bias and inconsistency of maximum likelihood using wrong models. Mol. Biol. Evol. 16:564-566.

Buckley, T. R. 2002. Model misspecification and probabilistic tests of topology: evidence from empirical data sets. Syst. Biol. 51:509-523.

Carroll, R. L. 1988. Vertebrate paleontology and evolution. Freeman, New York.

Carroll, R. L. 2000a. *Eocaecilia* and the origin of caecilians*in* Amphibian Biology (H. Heatwole, and R. L. Carroll, eds.). Surrey Beatty & Sons, Chipping Norton, Australia.

Carroll, R. L. 2000b. The Lissamphibian enigma. Pages 1270-1273 *in* Amphibian Biology (H. Heatwole, and R. L. Carroll, eds.). Surrey Beatty & Sons, Chipping Norton, Australia.

Carroll, R. L. 2001. The origin and early radiation of terrestrial vertebrates. J. Paleont. 75:1202-1213.

Carroll, R. L., C. Boisvert, J. Bolt, D. M. Green, N. Philip, C. Rolian, R. R. Schoch, and A. Tarenko. 2004. Changing patterns of ontogeny from osteolepiform fish through Permian tetrapods as a guide to the early evolution of land vertebrates. Pages 321-343 *in* Recent advances in the origin and early radiation of vertebrates (G. Arratia, M. V. H. Wilson, and R. Cloutier, eds.). Pfeil, München.

Chatterjee, S., and C. R. Scotese. 1999. The breakup of Gondwana and the evolution and biogeography of the Indian plate. Proc. Indian Natl. Sci. Acad. 65A:397-425.

Clary, D. O., and D. R. Wolstenholme. 1985. The mitochondrial DNA molecule of *Drosophila yakuba*: nucleotide sequence, gene organization and genetic code. J. Mol. Evol. 22:252-271.

Clayton, D. A. 1982. Replication of animal mitochondrial DNA. Cell 28:693-705.

Cummings, M. P., and A. Meyer. 2005. Magic bullets and golden rules: data sampling in molecular phylogenetics. Zoology 108:329-336.

Cummings, M. P., S. P. Otto, and J. Wakeley. 1995. Sampling properties of DNA sequence data in phylogenetic analysis. Mol. Biol. Evol. 12:814-822.

Cunningham, C. W., H. Zhu, and D. M. Hillis. 1998. Best-fit maximum-likelihood models for phylogenetic inference: empirical tests with known phylogenies. Evolution 52:978-987.

Doda, J. N., C. T. Wright, and D. A. Clayton. 1981. Elongation of displacement loop strands in human and mouse mitochondrial DNA is arrested near specific template sequences. Proc. Natl. Acad. Sci. USA 78:6116-6120.

Dorner, M., M. Altmann, S. Pääbo, and M. Morl. 2001. Evidence for import of a lysyl-tRNA into marsupial mitochondria. Mol. Biol. Cell. 12:2688-2698.

Duellman, W. E. 1975. On the classification of frogs. Occ. Pap. Mus. Nat. Hist. Univ. Kansas. 42:1-14.

Duellman, W. E., and L. Trueb. 1994. Biology of Amphibians. Johns Hopkins University Press, Baltimore, MD.

Efron, B. 1985. Bootstrap confidence intervals for a class of parametric problems. Biometrika 72:45-58.

Erixon, P., B. Svennblad, T. Britton, and B. Oxelman. 2003. Reliability of Bayesian posterior probabilities and bootstrap frequencies in phylogenetics. Syst. Biol. 52:665-673.

Estes, R., and M. H. Wake. 1972. The first fossil record of caecilian amphibians. Nature 239:228-231.

Evans, S. E., and M. Borsuk-Bialynicka. 1998. A stem-group frog from the Early Triassic of Poland. Acta Palaeo. Polonica 43:573-580.

Evans, S. E., A. R. Milner, and C. Werner. 1996. Sirenid salamanders and a gymnophionan amphibian from the Cretaceous of the Sudan. Palaeontology 39:77-95.

Evans, S. E., and D. Sigogneau-Russel. 2001. A stem-group caecilian (Lissamphibia: Gymnophiona) from the Lower Cretaceous of North Africa. Palaeontology 44:259-273.

Farris, J. S. 1983. The logical basis of phylogenetic systematics. Pages 7-36 *in* Advances in Cladistics (N. I. Platnick, and V. A. Funk, eds.). Columbia University Press, New York.

Feller, A. E., and S. B. Hedges. 1998. Molecular evidence for the early history of living amphibians. Mol. Phylogenet. Evol. 9:509-516.

Felsenstein, J. 1978. Cases in which parsimony or compatibility methods will be positively misleading. Syst. Zool. 27:401-410.

Felsenstein, J. 1981. Evolutionary trees from DNA sequences: a maximum likelihood approach. J. Mol. Evol. 17:368-376.

Felsenstein, J. 1985. Confidence limits on phylogenies: an approach using the bootstrap. Evolution 39:783-791.

Felsenstein, J. 2004. Inferring phylogenies. Sinauer Associates, Inc., Sunderland, MA.

Fitch, W. M. 1971. Toward defining the course of evolution: minimal change for a specific tree topology. Syst. Zool. 20:406-416.

Ford, L. S., and D. C. Cannatella. 1993. The major clades of frogs. Herpetol. Monogr. 7:94-117.

Frost, D. R. 2004. Amphibian species of the World: an online reference. V3.0 (22 August, 2004). http://research.amnh.org/herpetology/amphibia/index.html American Museum of Natural History, New York, NY.

Frost, D. R., T. Grant, J. Faivovich, R. H. Bain, A. Haas, C. F. B. Haddad, R. O. de Sá, A. Channing, M. Wilkinson, S. C. Donnellan, C. J. Raxworthy, J. A. Campbell, B. L. Blotto, P. Moler, R. C. Drewes, R. A. Nussbaum, J. D. Lynch, D. M. Green, and W. C. Wheeler. 2006. The amphibian tree of life. Bull. Am. Mus. Nat. His. 297:1-370.

Gao, K. Q., and N. H. Shubin. 2001. Late Jurassic salamanders from Northern China. Nature 410:574-577.

Goldman, N. 1990. Maximum likelihood inference of phylogenetic trees, with special reference to a Poisson process model of DNA substitution and to parsimony analysis. Syst. Zool. 39.

Goldman, N. 1993. Statistical tests of models of DNA substitution. J. Mol. Evol. 36:182–198.

Goldman, N. 1998a. Effects of sequence alignment procedures on estimates of phylogeny. BioEssays 20:287-290.

Goldman, N. 1998b. Phylogenetic information and experimental design in molecular systematics. Proc. R. Soc. Lond. B 265:1779-1786.

Goldman, N., J. P. Anderson, and A. G. Rodrigo. 2000. Likelihood-based tests of topologies in phylogenetics. Syst. Biol. 49:652-670.

Goldman, N., and S. Whelan. 2000. Statistical tests of gamma-distributed rate heterogeneity in models of sequence evolution in phylogenetics. Mol. Biol. Evol. 17:975-978.

Gower, D. J., A. Kupfer, O. V. Oommen, W. Himstedt, R. A. Nussbaum, S. P. Loader, B. Presswell, H. Müller, S. B. Krishna, R. Boistel, and M. Wilkinson. 2002. A molecular phylogeny of ichthyophiid caecilians (Amphibia: Gymnophiona: Ichthyophiidae): out of India or out of South East Asia? Proc. R. Soc. Lond. B 269:1563-1569.

Gower, D. J., S. P. Loader, C. B. Moncrieff, and M. Wilkinson. 2004. Niche separation and comparative abundance of *Boulengerula boulengeri* and *Scolecomorphus vittatus*

(Amphibia: Gymnophiona) in an East Usambara forest, Tanzania. Afr. J. Herpetol. 53:183-190.

Groth, J. G., and G. F. Barrowclough. 1999. Basal divergences in birds and the phylogenetic utility of the nuclear RAG-1 gene. Mol. Phylogenet. Evol. 12:115-123.

Gurnis, M. 1988. Large-scale mantle convection and the aggregation and dispersal of supercontinents. Nature 322:695–699.

Haas, A. 2003. Phylogeny of frogs as inferred from primarily larval characters (Amphibia: Anura). Cladistics 19:23-89.

Hasegawa, M., H. Kishino, and T. Yano. 1985. Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. J. Mol. Evol. 22:160-174.

Hay, J. M., I. Ruvinsky, S. B. Hedges, and L. R. Maxson. 1995. Phylogenetic relationships of amphibian families inferred from DNA sequences of mitochondrial 12S and 16S ribosomal RNA genes. Mol. Biol. Evol. 12:928-937.

Heatwole, H., and R. L. Carroll (eds) 2000. Amphibian Biology. Volume 4: Palaeontology. Surrey Beatty & Sons, Chipping Norton, Australia.

Hedges, S. B. 1992. The number of replications needed for accurate estimation of the bootstrap P value in phylogenetic studies. Mol. Biol. Evol. 9:366-369.

Hedges, S. B., and L. R. Maxson. 1993. A molecular perspective on lissamphibian phylogeny. Herpetol. Monogr. 7:27-42.

Hedges, S. B., K. D. Moberg, and L. R. Maxson. 1990. Tetrapod phylogeny inferred from 18S and 28S ribosomal RNA sequences and a review of the evidence for amniote relationships. Mol. Biol. Evol. 7:607-633.

Hedges, S. B., R. A. Nussbaum, and L. R. Maxson. 1993. Caecilian phylogeny and biogeography inferred from mitochondrial DNA sequences of the 12SrRNA and 16S rRNA genes (Amphibia: Gymnophiona). Herpetol. Monogr. 7:64-76.

Hennig, W. 1950. Grundzüge einer theorie der phylogenetischen systematik. Deutsche Zentral Verlag, Berlin.

Hillis, D. M. 1991. The phylogeny of amphibians: current knowledge and the role of cytogenetics. Pages 7-31 *in* Amphibian cytogenetics and evolution (S. K. Sessions, and D. M. Green, eds.). Academic Press, San Diego, CA.

Hillis, D. M., L. K. Ammerman, M. T. Dixon, and R. O. de Sá. 1993. Ribosomal DNA and the phylogeny of frogs. Herpetol. Monogr. 7:118-131.

Hillis, D. M., and J. J. Bull. 1993. An empirical test of bootstrapping as a method for assessing confidence in phylogenetic analysis. Syst. Biol. 42:182-192.

Hillis, D. M., C. Moritz, and B. K. Mable (eds) 1996. Molecular systematics. Sinauer Associates, Inc., Sunderland, MA.

Hillis, D. M., and J. J. Wiens. 2000. Molecules versus morphology in systematics: conflicts, artifacts, and misconceptions. Pages 1-19 *in* Phylogenetic analysis of morphological data (J. J. Wiens, ed.) Smithsonian Institution Press, Washington, DC.

Himstedt, W. 1996. Die Blindwühlen. Westarp Wissenschaften, Magdeburg.

Hixson, J. E., T. W. Wong, and D. A. Clayton. 1986. Both the conserved stem-loop and divergent 5'-flanking sequences are required for initiation at the human mitochondrial origin of light-strand DNA replication. J. Biol. Chem. 261:2384-2390.

Hoegg, S., M. Vences, H. Brinkmann, and A. Meyer. 2004. Phylogeny and comparative substitution rates of frogs inferred from sequences of three nuclear genes. Mol. Biol. Evol. 21:1188-1200.

Holder, M., and P. O. Lewis. 2003. Phylogeny estimation: traditional and Bayesian approaches. Nat. Rev. Genet. 4:275-284.

Huelsenbeck, J. P. 1997. Is Felsenstein zone a fly trap? Syst. Biol. 46:69-74.

Huelsenbeck, J. P., and K. A. Crandall. 1997. Phylogeny estimation and hypothesis testing using maximum likelihood. Ann. Rev. Ecol. Syst. 28:437-466.

Huelsenbeck, J. P., and D. M. Hillis. 1993. Success of phylogenetic methods in the four-taxon case. Syst. Biol. 42:247-264.

Huelsenbeck, J. P., D. M. Hillis, and R. Jones. 1996. Parametric bootstrapping in molecular phylogenetics: Applications and performance. Pages 19-45 *in* Molecular Zoology: Advances, Strategies, and Protocols (J. D. Ferarris, and S. R. Palumbi, eds.). Wiley-Liss, New York.

Huelsenbeck, J. P., and B. Rannala. 2004. Frequentist properties of Bayesian posterior probabilities of phylogenetic trees under simple and complex substitution models. Syst. Biol. 53:904-913.

Huelsenbeck, J. P., F. R. Ronquist, R. Nielsen, and J. P. Bollback. 2001. Bayesian inference of phylogeny and its impact on evolutionary biology. Science 294:2310-2314.

Irwin, D. M., T. D. Kocher, and A. C. Wilson. 1991. Evolution of the cytochrome *b* gene of mammals. J. Mol. Evol. 32:128-144.

Jameson, D., A. P. Gibson, C. Hudelot, and P. G. Higgs. 2003. OGRe: a relational database for comparative analyses of mitochondrial genomes. Nucleic Acids Res. 31:202-206.

Janke, A., O. Magnell, G. Wieczorek, M. Westerman, and U. Arnason. 2002. Phylogenetic analyses of 18S rRNA and the mitochondrial genomes of the Wombat, *Vombatus*

ursinus, and the spiny anteater, *Tachyglossus aculeatus*: increased support for the Marsupionta hypothesis. J. Mol. Evol. 54:71-80.

Janke, A., X. Xu, and U. Arnason. 1997. The complete mitochondrial genome of the wallaroo (*Macropus robustus*) and the phylogenetic relationships among Monotremata, Marsupialia, and Eutheria. Proc. Natl. Acad. Sci. USA 94:1276-1281.

Jenkins, F. A., and D. M. Walsh. 1993. An early Jurassic caecilians with limbs. Nature 365:246-249.

Johnson, K. P., and M. D. Sorenson. 1998. Comparing Molecular Evolution in Two Mitochondrial Protein Coding Genes (Cytochrome b and ND2) in the Dabbling Ducks (Tribe: Anatini). Mol. Phylogenet. Evol. 10:82-94.

Jones, D. T., W. R. Taylor, and J. M. Thornton. 1992. The rapid generation of mutation data matrices from protein sequences. Comp. Appl. Biosci. 8:275-282.

Kimura, M. 1968. Evolutionary rate at the molecular level. Nature 217:624-626.

Kimura, M. 1983. The neutral theory of molecular evolution. Cambridge University Press, Cambridge.

Kishino, H., and M. Hasegawa. 1989. Evaluation of the maximum likelihood estimate of the evolutionary tree topologies from DNA sequence data, and the branching order in Hominoidea. J. Mol. Evol. 29:170-179.

Kishino, H., J. L. Thorne, and W. J. Bruno. 2001. Performance of a divergence time estimation method under a probabilistic model of rate evolution. Mol. Biol. Evol. 18:352-361.

Kocher, T. D., W. K. Thomas, A. Meyer, S. V. Edwards, S. Pääbo, F. X. Villablanca, and A. C. Wilson. 1989. Dynamics of mitochondrial DNA evolution in animals: amplification and sequencing with conserved primers. Proc. Natl. Acad. Sci. USA 86:6196-6200.

Kumazawa, Y., and M. Nishida. 1993. Sequence evolution of mitochondrial tRNA genes and deep-branch animal phylogenetics. J. Mol. Evol. 37:380-398.

Kumazawa, Y., H. Ota, M. Nishida, and T. Ozawa. 1996. Gene rearrangements in snake mitochondrial genomes: highly concerted evolution of control-region-like sequences duplicated and inserted into a tRNA cluster. Mol. Biol. Evol. 13:1242-1254.

Kupfer, A., J. Nabhitabhata, and W. Himstedt. 2004. Reproductive ecology of female caecilian amphibians (genus *Ichthyophis*): a baseline study. Biol. J. Linn. Soc. 83:207–-217.

Larson, A. 1991. A molecular perspective on the evolutionary relationships of the salamander families. Evol. Biol. 25:211-277.

Larson, A., and W. W. Dimmick. 1993. Phylogenetic relationships of the salamander families: an analysis of congruence among morphological and molecular characters. Herpetol. Monogr. 7:77-93.

Larson, A., and A. C. Wilson. 1989. Patterns of ribosomal RNA evolution in salamanders. Mol. Biol. Evol. 6:131-154.

Laurent, R. 1979. Esquisse d'une phylogenèse des anoures. Bull. Soc. Zool. France 104:397-422.

Laurin, M. 1998. The importance of global parsimony and historical bias in understanding tetrapod evolution. Part I. Systematics, middle ear evolution and jaw suspension. Ann. Sci. Nat. Zool. Biol. Anim. 13eme Série 19:1-42.

Laurin, M., and R. Reisz. 1997. A new perspective on tetrapod phylogeny. Pages 9-59 *in* Amniote origins - Completing the transition to land (S. S. Sumida, and K. L. Martin, eds.). Academic Press, New York.

Lee, M. S. Y., and J. S. Anderson. 2006. Molecular clocks and the origin(s) of modern amphibians. Mol. Phylogenet. Evol. 40:635-639.

Li, W.-H., and D. Graur. 1991. Fundamentals of Molecular Evolution. Sinauer, Sunderland, MA.

Loader, S. P., M. Wilkinson, D. J. Gower, and C. A. Msuya. 2003. A remarkable young *Scolecomorphus vittatus* (Amphibia: Gymnophiona: Scolecomorphidae) from the North Pare Mountains, Tanzania. J. Zool. 259:93-101.

Lopez, J. V., M. Culver, J. C. Stephens, W. E. Johnson, and S. J. O'Brien. 1997. Rates of nuclear and cytoplasmic mitochondrial DNA sequence divergence in mammals. Mol. Biol. Evol. 14:277-286.

Lunt, D. H., and B. C. Hyman. 1997. Animal mitochondrial DNA recombination. Nature 387:247.

Macey, J. R., A. Larson, N. B. Ananjeva, Z. Fang, and T. J. Papenfuss. 1997. Two novel gene orders and the role of light-strand replication in rearrengement of the vertebrate mitochondrial genome. Mol. Biol. Evol. 14:91-104.

Maley, L. E., and C. R. Marshall. 1998. The coming of age of molecular systematics. Science 279:505-506.

Martin, A. P. 1999. Susbtitution rates of organelle and nuclear genes in sharks: implicating metabolic rate (again). Mol. Biol. Evol. 16:996-1002.

Martin, A. P., and S. R. Palumbi. 1993. Body size, metabolic rate, generation time and the molecular clock. Proc. Natl. Acad. Sci. USA 90:4087-4091.

McGowan, G., and S. E. Evans. 1995. Albanerpetontid amphibians from the Cretaceous of Spain. Nature 373:143-145.

Meyer, A. 1993. Evolution of mitochondrial DNA in fishes. Pages 1-38 *in* Biochemistry and molecular biology of fishes (P. W. Hochachka, and T. P. Mommsen, eds.). Elsevier Science.

Milner, A. R. 1988. The relationships and origin of living amphibians. Pages 59-102 *in* The phylogeny and classification of the tetrapods (M. J. Benton, ed.) Clarendon Press, Oxford.

Milner, A. R. 1993. The Paleozoic relatives of lissamphibians. Herpetol. Monogr. 7:8-27.

Mindell, D. P., and R. L. Honeycutt. 1990. Ribosomal RNA in vertebrates: evolution and phylogenetic applications. Ann. Rev. Ecol. Syst. 21:541-566.

Mindell, D. P., M. D. Sorenson, D. E. Dimcheff, M. Hasegawa, J. C. Ast, and T. Yuri. 1999. Interordinal relationships of birds and other reptiles based on whole mitochondrial genomes. Syst. Biol. 48:138-152.

Miya, M., and M. Nishida. 2000. Use of mitogenomic information in teleostean molecular phylogenetics: a tree-based exploration under the maximum-parsimony optimality criterion. Mol. Phylogenet. Evol. 17:437-455.

Miya, M., H. Takeshima, H. Endo, N. B. Ishiguro, J. G. Inoue, T. Mukai, T. P. Satoh, M. Yamaguchi, A. Kawaguchi, K. Mabuchi, S. M. Shirai, and M. Nishida. 2003. Major patterns of higher teleostean phylogenies: a new perspective based on 100 complete mitochondrial sequences. Mol. Phylogenet. Evol. 26:121-138.

Moritz, C., and W. M. Brown. 1986. Tandem duplications of D-loop and ribosomal RNA sequences in lizard mitochondrial DNA. Science 233:1425-1427.

Moritz, C., and W. M. Brown. 1987. Tandem duplications in animal mitochondrial DNAs: variation in incidence and gene content among lizards. Proc. Natl. Acad. Sci. USA 84:7183-7187.

Moritz, C., T. E. Dowling, and W. M. Brown. 1987. Evolution of animal mitochondrial DNA: relevance for population biology and systematics. Annu. Rev. Ecol. Syst. 18:269-292.

Mueller, R. L. 2006. Evolutionary rates, divergence dates, and the performance of mitochondrial genes in Bayesian phylogenetic analysis. Syst. Biol. 55:289-300.

Murphy, W. J., E. Eizirik, S. J. O'Brien, O. Madsen, M. DScally, C. J. Douady, E. teeling, O. A. Ryder, M. J. Stanhope, W. W. de Jong, and M. S. Springer. 2001. Resolution of the early palcental mammal radiation using bayesian phylogenetics. Science 294:2348-2350.

Nei, M. 1987. Molecular evolutionary genetics. Columbia University Press, New York.

Nei, M., and S. Kumar. 2000. Molecular evolution and phylogenetics. Oxford University Press, Oxford.

Noble, G. K. 1931. The biology of the Amphibia. McGraw-Hill, New York.

Nussbaum, R. A. 1977. Rhinatrematidae: a new family of caecilians (Amphibia: Gymnophiona). Occ. Pap. Mus. Zool. Univ. Michigan 682:1-30.

Nussbaum, R. A. 1979. The taxonomic status of the caecilian genus *Uraeotyphlus* Peters. Occ. Pap. Mus. Zool. Univ. Michigan 687:1-20.

Nussbaum, R. A. 1985. Systematics of the caecilians (Amphibia: Gymnophiona) of the family Scolecomorphidae. Occ. Pap. Mus. Zool. Univ. Michigan 713:1-49.

Nussbaum, R. A. 1991. Cytotaxonomy of caecilians. Pages 22-76 *in* Amphibian cytogenetics and evolution (S. K. Sessions, and D. M. Green, eds.). Academic Press, San Diego, CA.

Nussbaum, R. A. 1998. Caecilians. Pages 52-59 *in* Encyclopedia of reptiles and amphibians (H. G. Cogger, and R. G. Zweifel, eds.). Academic Press, San Diego.

Nussbaum, R. A., and M. Wilkinson. 1989. On the classification and phylogeny of caecilians (Amphibia: Gymnophiona), a critical review. Herpetol. Monogr. 3:1-42.

Ohta, T. 2002. Near-neutrality in evolution of genes and in gene regulation. Proc. Natl. Acad. Sci. USA 99:16134-16137.

Ota, R., and D. Penny. 2003. Estimating changes in mutational mechanisms of evolution. J. Mol. Evol. 57:S233-S240.

Ota, R., P. J. Waddell, M. Hasegawa, H. Shimodaira, and H. Kishino. 2000. Appropiate likelihood ratio test and marginal distributions for evolutionary tree models with constraints on parameters. Mol. Biol. Evol. 17:798-803.

Pääbo, S., H. Poinar, D. Serre, V. Jaenicke-Despres, J. Hebler, N. Rohland, M. Kuch, J. Krause, L. Vigilant, and M. Hofreiter. 2004. Paabo S, Poinar H, Serre D, Jaenicke-Despres V, Hebler J, et al. (2004) Genetic analyses from ancient DNA. Annu Rev Genet 38: 645–679. Annu. Rev. Genet. 38:645–679.

Pääbo, S., W. K. Thomas, K. M. Whitfield, and Y. Kumazawa. 1991. Rearrangements of mitochondrial transfer RNA genes in marsupials. J. Mol. Evol. 33:426-430.

Page, R. D. M., and E. C. Holmes. 1998. Molecular evolution: a phylogenetic approach. Blackwell Science, Oxford.

Palumbi, S. R., A. Martin, S. Romano, W. Owen MacMillan, L. Stice, and G. Grabowski. 1991. The simple fool's guide to PCR. Department of Zoology, University of Hawaii, Honolulu.

Panchen, A. L., and T. R. Smithson. 1987. Character diagnosis, fossils and the origin of tetrapods. Biol. Rev. 62:341-438.

Parsons, T. S., and E. E. Williams. 1963. The relationships of the modern Amphibia: a re-examination. Q. Rev. Biol. 38:26-53.

Phillips, A., D. Janies, and W. Wheeler. 2000. Multiple Sequence Alignment in Phylogenetic Analysis. Mol. Phylogenet. Evol. 16:317-330.

Pitman III, W. C., S. Cande, J. LaBrecque, and J. Pindell. 1993. Fragmentation of Gondwana: the separation of Africa from South America. Pages 15-34 *in* Biological relationships between Africa and South America (P. Goldblatt, ed.) Yale University Press, New Haven, CT.

Posada, D. 2003. Selecting models of evolution. Pages 256-282 *in* The phylogenetic handbook (M. Salemi, and A.-M. Vandamme, eds.). Cambridge University Press, Cambridge.

Posada, D., and K. A. Crandall. 2001. Selecting the best-fit model of nucleotide substitution. Syst. Biol. 50:580-601.

Pugener, L. A., A. M. Maglia, and L. Trueb. 2003. Revisiting the contribution of larval characters to an analysis of phylogenetic relationships of basal anurans. Zool. J. Linn. Soc. 139:129-155.

Rabinowitz, P. D., M. F. Coffin, and D. Falvey. 1983. The separation of Madagascar and Africa. Science 220:67-69.

Rage, J., and Z. Rocek. 1989. Redescription of *Triadobatrachus massinoti* (Piveteau, 1936) an anuran amphibian from the early Triassic. Paleontographica Abt. A 206:1-16.

Rage, J. C. 1986. Le plus ancien Amphibien apode (Gymnophiona) fossile. Remarques sur la répartition et l'historie paleobiogeographique des Gymnophiones. Comtes rendus de l'Académie des Sciences, Paris 302:1033-1036.

Rage, J. C. 1991. Gymnophionan amphibian from Early Paleocene (Santa Lucía Formation) of Tiupampa: the earliest known Gymnophiona. Pages 499-501 *in* Fósiles y facies de Bolivia, 1. Vertebrados. Revista técnica de yacimientos petrolíferos fiscales Bolivianos. 12: 359-718 (R. Suarez-Soruco, ed.).

Rage, J. C., and P. Janvier. 1982. Le problème de la monophylie des amphibiens actuels, à la lumière des nouvelles donnees sur les affinités des tétapodes. Geobios 6:65-83.

Raimond, R., I. Marcade, D. Bouchon, T. Rigaud, J.-P. Borry, and C. Souty-Grosset. 1999. Organization of the large mitochondrial genome in the isopod *Armadillidium vulgare*. Genetics 151:203–210.

Rannala, B., and Z. Yang. 1996. Probability distribution of molecular evolutionary trees: a new method of phylogenetic inference. J. Mol. Evol. 43:304-311.

Reeves, J. H. 1992. Heterogeneity in the substitution process of amino acid sites of proteins coded for by mitochondrial DNA. J. Mol. Evol. 35:17-31.

Rest, J. S., J. C. Ast, C. C. Austin, P. J. Waddell, E. A. Tibbetts, J. M. Hay, and D. P. Mindell. 2003. Molecular systematics of primary reptilian lineages abd the tuatara mitochondrial genome. Mol. Phylogenet. Evol. 29:289-297.

Rodríguez, F., J. F. Oliver, A. Marín, and J. R. Medina. 1990. The general stochastic model of nucleotide substitution. J. Theor. Biol. 142:485-501.

Roe, B. A., M. Din-Pow, R. K. Wilson, and J. F. Wong. 1985. The complete nucleotide sequence of the *Xenopus laevis* mitochondrial genome. J. Biol. Chem. 260:9759-9774.

Roelants, K., D. J. Gower, M. Wilkinson, S. P. Loader, S. D. Biju, K. Guillaume, and F. Bossuyt. High extinction rates and fluctuating diversification in the history of modern amphibians. Unpublished manuscript.

Rokas, A., and P. W. H. Holland. 2000. Rare genomic changes as a tool for phylogenetics. Trends Ecol. Evol. 15:454-459.

Russo, C. A. M., N. Takezaki, and M. Nei. 1996. Efficiencies of different genes and different tree-building methods in recovering a known vertebrate phylogeny. Mol. Biol. Evol. 13:525-536.

Ruta, M., M. I. Coates, and D. L. J. Quicke. 2003. Early tetrapod relationships revisited. Biol. Rev. 78:251–345.

Rzhetsky, A., and M. Nei. 1992. A simple method for estimating and testing minimum-evolution trees. Mol. Biol. Evol. 9:945-967.

Saiki, R. K., D. H. Gelfand, S. Stoffel, S. Scharf, R. Higuchi, G. T. Horn, K. B. Mullis, and H. A. Erlich. 1988. Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. Science 239:487-491.

Saitou, N., and M. Nei. 1987. The neighbor-joining method: A new method for reconstructing phylogenetic trees. Mol. Biol. Evol. 4:406-425.

San Mauro, D., M. García-París, and R. Zardoya. 2004. Phylogenetic relationships of discoglossid frogs (Amphibia: Anura: Discoglossidae) based on complete mitochondrial genomes and nuclear genes. Gene 343:357-366.

Sanderson, M. J. 1997. A nonparametric approach to estimating divergence times in the absence of rate constancy. Mol. Biol. Evol. 14:1218-1231.

Sanderson, M. J. 2002. Estimating absolute rates of molecular evolution and divergence times: a penalized likelihood approach. Mol. Biol. Evol. 19:101-109.

Sanderson, M. J. 2003. R8S: inferring absolute rates of molecular evolution and divergence times in the absence of a molecular clock. Bioinformatics 19:301-302.

Schatz, D. G., M. A. Oettinger, and D. Baltimore. 1989. The V(D)J Recombination Activating Gene, RAG-1. Cell 59:1035-1048.

Schoch, R. R., and A. R. Milner. 2004. Structure and implications of theories on the origin of lissamphibians. Pages 345-377 *in* Recent advances in the origin and early radiation of vertebrates (G. Arratia, M. V. H. Wilson, and R. Cloutier, eds.). Pfeil, München.

Seutin, G., B. F. Lang, D. P. Mindell, and R. Morais. 1994. Evolution of the WANCY region in amniote mitochondrial DNA. Mol. Biol. Evol. 11:329-340.

Shadel, G. S., and D. A. Clayton. 1997. Mitochondrial DNA maintenance in vertebrates. Annu. Rev. Biochem. 66:409-435.

Shimodaira, H. 2002. An approximately unbiased test of phylogenetic tree selection. Syst. Biol. 51:492-508.

Shimodaira, H., and M. Hasegawa. 1999. Multiple comparisons of Log-likelihoods with applications to phylogenetic inference. Mol. Biol. Evol. 16:1114-1116.

Smith, A. G., D. G. Smith, and B. M. Funnell. 1994. Atlas of Mesozoic and Cenozoic coastlines. Cambridge University Press, Cambridge.

Springer, M. S., R. W. DeBry, C. J. Douady, H. M. Amrine, O. Madsen, W. W. deJong, and M. J. Stanhope. 2001. Mitochondrial versus nuclear gene sequences in deep-level mammalian phylogeny reconstruction. Mol. Biol. Evol. 18:132-143.

Stevens, J. R., and C. J. Schofield. 2003. Phylogenetics and sequence analysis - some problems for the unwary. Trends Parasitol. 19:582-588.

Strimmer, K., and A. Rambaut. 2001. Inferring confidence sets of possible misspecified gene trees. Proc. R. Soc. London B 269:137-142.

Suzuki, Y., G. V. Glazko, and M. Nei. 2002. Overcredibility of molecular phylogenies obtained by Bayesian phylogenetics. Proc. Natl. Acad. Sci. USA 99:16138–16143.

Swofford, D. L., G. J. Olse, P. J. Waddell, and D. M. Hillis. 1996. Phylogenetic inference. Pages 407-514 *in* Molecular systematics (D. M. Hillis, C. Moritz, and B. K. Mable, eds.). Sinnauer Associates, Inc., Sunderland, MA.

Taylor, E. H. 1968. The Caecilians of the world: A taxonomic analysis. University of Kansas Press, Lawrence, KS.

Taylor, E. H. 1969. A new family of African Gymnophiona. Univ. Kansas Sci. Bull. 48:297-305.

Templeton, A. R. 1983. Phylogenetic inference from restriction endonuclease cleavage site maps with particular reference to the evolution of human and the apes. Evolution 37:221-244.

Thompson, C. B. 1995. New insights into V(D)J recombination and its role in the evolution of the immune system. Immunity 3:531–539.

Thompson, J. D., F. Plewniak, and O. Poch. 1999. A comprehensive comparison of multiple sequence alignment programs. Nucleic Acids Res. 7:2682-2690.

Thorne, J. L., and H. Kishino. 2002. Divergence time and evolutionary rate estimation with multilocus data. Syst. Biol. 51:689-702.

Thorne, J. L., H. Kishino, and I. S. Painter. 1998. Estimating the rate of evolution of the rate of molecular evolution. Mol. Biol. Evol. 15:1647-1657.

Townsend, T. M., A. Larson, E. Louis, and J. R. Macey. 2004. Molecular Phylogenetics of Squamata: The Position of Snakes, Amphisbaenians, and Dibamids, and the Root of the Squamate Tree. Syst. Biol. 53:735-757.

Trueb, L., and R. Cloutier. 1991. A phylogenetic investigation of the inter- and intrarelationships of the Lissamphibia (Amphibia: Temnospondyli). Pages 223-313 *in* Origins of the major groups of tetrapods: Controversies and consensus (H. P. Schultze, and L. Trueb, eds.). Cornell Univ. Press, Ithaca, NY.

Vences, M., D. R. Vieites, F. Glaw, H. Brinkmann, J. Kosuch, M. Veith, and A. Meyer. 2003. Multiple overseas dispersal in amphibians. Proc. R. Soc. Lond. B 270:2435-2442.

Wake, M. H. 1977. The reproductive biology of caecilians: an evolutionary perspective. Pages 73-101 *in* Reproductive biology of amphibians (E. H. Taylor, and S. I. Guttman, eds.). Plenum Press, New York.

Walberg, M. W., and D. A. Clayton. 1981. Sequence and properties of the human KB cell and mouse L cell D-Loop regions of mitochondrial DNA. Nucleic Acids Res. 9: 5411-5421.

Werner, C. 1994. Der erste Nachweis von Gymnophionen (Amphibia) in der Kreide (Wadi-Milk-Formation, Sudan). Neues jahrbuch für geologie und paläontologie, Monatshefte 1994:633-640.

Whelan, S., P. Liò, and N. Goldman. 2001. Molecular phylogentics: state-of-the-art methods for looking into the past. Trends Genet. 17:262-272.

Wilkinson, M. 1992. The phylogenetic position of the Rhinatrematidae (Amphibia: Gymnophiona): evidence from the larval lateral line system. Amphibia-Reptilia 13:74-79.

Wilkinson, M. 1996. The heart and aortic arches of rhinatrematid caecilians (Amphibia: Gymnophiona). Zoomorphology 105:277-295.

Wilkinson, M. 1997. Characters, congruence and quality: a study of neuroanatomical and traditional data in caecilian phylogeny. Biol. Rev. 72:423-470.

Wilkinson, M., S. P. Loader, D. J. Gower, J. A. Sheps, and B. L. Cohen. 2003. Phylogenetic relationships of African caecilians (Amphibia: Gymnophiona): insights from mitochondrial rRNA gene sequences. Afr. J. Herpetol. 52:83-92.

Wilkinson, M., and R. A. Nussbaum. 1992. Taxonomic status of *Pseudosiphonops ptychodermis* Taylor and *Mimosiphonops vermiculatus* Taylor (Amphibia: Gymnophiona: Caeciliaidae) with a description of a new species. J. Nat. Hist. 26:675-688.

Wilkinson, M., and R. A. Nussbaum. 1996. On the phylogenetic position of the Uraeotyphlidae (Amphibia: Gymnophiona). Copeia 1996:550-562.

Wilkinson, M., and R. A. Nussbaum. 1997. Comparative morphology and evolution of the lungless caecilian *Atretochoana eiselti* (Taylor) (Amphibia: Gymnophiona: Typhlonectidae). Biol. J. Linn. Soc. 62:39-109.

Wilkinson, M., and R. A. Nussbaum. 1999. Evolutionary relationships of the lungless caecilian *Atretochoana eiselti* (Amphibia: Gymnophiona: Typhlonectidae). Zool. J. Linn. Soc. 126:191-223.

Wilkinson, M., and R. A. Nussbaum. 2006. Caecilian phylogeny and classification. Pages 39-78 *in* Reproductive biology and phylogeny of Gymnophiona (Caecilians) (J.-M. Exbrayat, ed.) Science Publishers, Enfield, NH.

Wilkinson, M., J. A. Sheps, O. V. Oommen, and B. L. Cohen. 2002. Phylogenetic relationships of Indian caecilians (Amphibia: Gymnophiona) inferred from mitochondrial rRNA gene sequences. Mol. Phylogenet. Evol. 23:401-407.

Wolstenholme, D. R. 1992. Animal mitochondrial DNA: structure and evolution. Int. Rev. Cytol. 141:173-216.

Wong, T. W., and D. A. Clayton. 1985. In vitro replication of human mitochondrial DNA: accurate initiation at the origin of light-strand synthesis. Cell 42: 951-958.

Wu, C. I., and W. H. Li. 1985. Evidence for higher rates of nucleotide substitution in rodents than in man. Proc. Natl. Acad. Sci. USA 82:1741–1745.

Yang, Z. 1994a. Estimating the pattern of of nucleotide substitution. J. Mol. Evol. 39:105-111.

Yang, Z. 1994b. Maximum likelihood phylogenetic estimation from DNA sequences with variable rates over sites: approximate methods. J. Mol. Evol. 39:306-314.

Zardoya, R., E. Malaga-Trillo, M. Veith, and A. Meyer. 2003. Complete nucleotide sequence of the mitochondrial genome of a salamander, *Mertensiella luschani*. Gene 317:17-27.

Zardoya, R., and A. Meyer. 1996a. The complete nucleotide sequence of the mitochondrial genome of the lungfish (*Protopterus dolloi*) supports its phylogenetic position as a close relative of land vertebrates. Genetics 142:1249-1263.

Zardoya, R., and A. Meyer. 1996b. Phylogenetic performance of mitochondrial protein-coding genes in resolving relationships among vertebrates. Mol. Biol. Evol. 13:933-942.

Zardoya, R., and A. Meyer. 1998. Complete mitochondrial genome suggests diapsid affinities of turtles. Proc. Natl. Acad. Sci. USA 95:14226-14231.

Zardoya, R., and A. Meyer. 2000. Mitochondrial evidence on the phylogenetic position of Caecilians (Amphibia: Gymnophiona). Genetics 155:765-775.

Zardoya, R., and A. Meyer. 2001. On the origin of and phylogenetic relationships among living amphibians. Proc. Natl. Acad. Sci. USA 98:7380-7383.

Zhang, P., Y. Q. Chen, Y. F. Liu, H. Zhou, and L. H. Qu. 2003a. The complete mitochondrial genome of the Chinese giant salamander, *Andrias davidianus* (Amphibia: Caudata). Gene 311:93-98.

Zhang, P., Y. Q. Chen, H. Zhou, X. L. Wang, and L. H. Qu. 2003b. The complete mitochondrial genome of a relic salamander, *Ranodon sibiricus* (Amphibia: Caudata) and implications for amphibian phylogeny. Mol. Phylogenet. Evol. 28:620-626.

Zhang, P., H. Zhou, Y.-Q. Chen, Y.-F. Liu, and L.-H. Qu. 2005. Mitogenomic perspectives on the origin and phylogeny of living amphibians. Syst. Biol. 54:391-400.

Zharkikh, A., and W.-H. Li. 1992. Statistical properties of bootstrap estimation of phylogenetic variability from nucleotide sequences. II. Four taxa without a molecular clock. J. Mol. Evol. 35:356-366.

Zuckerkandl, E., and L. Pauling. 1965. Evolutionary divergence and convergence in proteins. Pages 97-166 *in* Evolving genes and proteins (V. Bryson, and H. Vogel, eds.). Academic Press, New York.